DIAGNOSTIC ASSAYS AND KITS FOR BODY MASS AND CARDIOVASCULAR DISORDERS

Related Applications

This application is a continuation-in-part of U.S. Application No. 09/031,626, filed on February 27, 1998, which is a continuation-in-part of U.S. Application No. 08/890,979, filed on July 10, 1997 (U.S. Patent No. 6,030,778), the contents of which are incorporated herein in their entirety by this reference.

1. Background of the Invention

Coronary heart disease is a major health risk throughout the industrialized world. Atherosclerosis, the most prevalent of cardiovascular diseases, is the principal cause of heart attack, stroke, and gangrene of the extremities, and thereby the principle cause of death in the United States. Dyslipidemia is associated with the development of coronary heart disease (CHD) and atherosclerosis. Although historically much emphasis has been placed on total plasma cholesterol levels as a risk factor for coronary heart disease, it has been clearly established that low levels of high density lipoprotein cholesterol (HDL) is an independent risk factor for this disease. Family and twin studies have shown that there are genetic components that affect HDL levels. However, mutations in the main protein components of HDL (ApoA1 and ApoAII) and in the enzymes that are known to be involved in HDL metabolism (*e.g.*, CETP, HL, LPL and LCAT) do not explain all of the genetic factors affecting HDL levels in the general population (J. L. Breslow, in *The Metabolic and Molecular Bases of Inherited Disease,* C.R. Scriver, A.L. Beaudet, W. Sly, D. Valle, Eds. (McGraw-Hill, New York, 1995), pp 2031-2052; and S.M. Grundy, (1995) *J. Am. Med. Assoc.* 256: 2849). This finding in combination with the fact that the mechanisms of HDL metabolism are poorly understood, suggests that there are other as yet unknown factors that contribute to the genetic variability of lipid levels, including HDL levels.

Another disorder that is often associated with high triglyceride and low high density lipoprotein (HDL) concentrations is obesity, which renders a subject susceptible to cardiovascular diseases, such as ischemia, restenosis, congestive heart failure, and

atherosclerosis. Severely obese individuals (weighing 60% over a normal weight) have a high risk of developing cardiorespiratory problems. They are also at risk of developing chronic hypoventilation, which can lead to hypercapnia, pulmonary hypertension, and heart failure. Severe episodic hypoxia, which can cause arrhythmias and sudden death, is 10 times more common in the severely obese. Severely obese individuals are also at increased risk of suffering from obstructive sleep apnea, pickwickian syndrome (i.e., daytime hypoventilation, somnolence, polycythemia, cor pulmonale), and renal vein thrombosis. ("Cecil Essentials of Medicine", Andreoli et al., Third Edition, 1993, W.B. Saunders Company).

Moderate obesity (corresponding to a weight between 20-60% above normal weight) poses increased risk of early mortality. Obese individuals suffer more frequently than non obese individuals from hypertension. Type II diabetes mellitus can also be aggravated by excess weight. Obesity can also increase the risk of a subject developing cholelithiasis and endometrial carcinoma.

The risk of dyslipidemia and CHD is much greater in both non-insulin-dependent (type 2) diabetes mellitus (T2DM) patients and relatives of T2DM patients compared to non-diabetics (Groop L, et al. (1996) *Diabetes* 45(11):1585-93; Shaw JT, et al. (1999) *Diabetologia* 42(1):24-27. Among diabetics, the predominant dyslipidemia consists of low levels of high-density lipoprotein cholesterol (HDL) and high triglycerides (TG). (Knudsen P, et al. (1995) *Diabetologia* 38(3):344-50; Ginsberg HN (1996) *Diabetes* 45(suppl 3):S27-S31; Ginsberg HN (1991) *Diabetes Care*;14:839-855; Semenkovich DF, Heinecke JW (1997) *Diabetes* 46:327-34).

One candidate factor that is likely to be involved both in obesity and cardiovascular disease is the SR-BI receptor, which has been shown to bind HDL and LDL cholesterol and mediate uptake into cells (Acton, S. et al., (1996) *Science 271*:518-520). SR-BI is likely to contribute to genetic lipoprotein variability, thereby playing a role in the development of lipid metabolism disorders and thus generally, cardiovascular diseases.

In addition, cholesterol gallstone formation could be caused by a defective SR-BI receptor, since the SR-BI receptor is likely to be involved in transferring HDL cholesterol from extrahepatic tissues to the liver (reverse cholesterol transport) *e.g.* for incorporation into bile (J. L. Breslow, in *The Metabolic and Molecular Bases of Inherited Disease,* C.R. Scriver, A.L. Beaudet, W. Sly, D. Valle, Eds. (McGraw-Hill, New York, 1995), pp 2031-

2052; S.M. Grundy, (1995) *J. Am. Med. Assoc.* 256: 2849; G. Assman, A. von Eckardstein, H.B. Brewer Jr. in *The Metabolic and Molecular Bases of Inherited Disease,* C.R. Scriver, A.L. Beaudet, W. Sly, D. Valle, Eds. (McGraw-Hill, New York, 1995), pp 2053-2072; W.J. Johnson et al., (1991) *Biochem. Biophys. Acta 1085*:273; M.N. Pieters et al., (1994) *Ibid 1225*:125; and C. J. Fielding and P. E. Fielding, (1995) *J. Lipid Res 36*:211).

Further, a defective SR-BI receptor or abnormal levels of SR-BI receptor could influence the fertility of a subject, since SR-BI appears to be involved in HDL cholesteryl ester delivery to steroidogenic tissues (ovary, adrenal glands and testis) for hormone synthesis (Acton, S. et al., (1996) *Science 271*:518-520; Landschulz, et al., (1996) *J. Clin. Invest. 98*:984-95; J. M. Anderson and J. M. Dietschy (1981) *J. Biol. Chem. 256:* 7362; M.S. Brown et al., (1979) *Recent Prog Horm. Res. 35*:215; J.T. Gwynne and J.F. Strauss III, (1982) *Endocr. Rev. 3*:299; B.D. Murphy et al., (1985) *Endocrinology 116*: 1587).

The SR-BI receptor (Scavenger Receptor-BI) is a scavenger receptor that mediates endocytosis of unmodified and modified lipoproteins, *e.g.*, LDL, acetylated LDL, oxidized LDL (Acton et al. (1994) J. Biol. Chem. 269:21003), HDL ((Acton, S. et al., (1996) *Science 271*:518-520), anionic phospholipids (Rigotti et al. (1995) J. Biol. Chem. 270:16221), negatively charged liposomes and apoptotic cells (Fukasawa et al. (1996) Exp. Cell Res. 222:246). The human SR-BI receptor (also termed "CLA-1) exists in two differentially spliced forms (Calvo and Vega (1993) J. Biol. Chem. 268:18929). The predominant form of human SR-BI is a protein of 509 amino acids. The shorter form of the SR-BI receptor has 409 amino acids, and is lacking the 100 amino acids located 42 amino acids downstream of the initiation codon (Calvo and Vega, *supra*). The nucleotide sequence of a cDNA encoding human SR-BI is disclosed in Calvo and Vega, *supra* and the nucleotide sequence of a cDNA encoding hamster SR-BI is disclosed in Acton et al. (1994) J. Biol. Chem. 269:21003 and in PCT Application WO 96/00288.

2. Summary of the Invention

The present invention is based at least in part on the discovery of the genomic structure of the human SR-BI gene and on the identification of polymorphic regions within the gene, which are associated with specific diseases or disorders, including abnormal

body mass and abnormal lipoprotein levels, *i.e.*, low HDL and high LDL levels. Furthermore, the present invention is based, also in part, on the discovery of significant associations between two common polymorphisms within the SR-B1 gene and low levels in three ethnically diverse populations. Furthermore, the associations between SR-B1 polymorphisms and low HDL are influenced by gender, which implies an interaction with hormonal status. Therefore, these SR-B1 polymorphisms may be useful in predicting the effect of hormone replacement therapy (HRT) on HDL levels in female subjects, *e.g.*, postmenopausal female subjects.

The human SR-BI gene contains 12 coding exons and one non coding exon (exon 13). The structure of the gene and the position of the introns relative to the nucleotide sequence of the exons are shown in Figures 1, 2, and 3.

Several polymorphic regions that are associated with specific diseases or disorders, have been found in the human SR-BI gene by analyzing the DNA of a specific population of individuals. One polymorphism found in the population is a change from a guanine to an adenine at position 146 in exon 1, which results in a change from a glycine to a serine at amino acid residue 2 of the encoded protein. A second polymorphism is a change from a guanine to an adenine at position 119 in exon 3, which results in a change from a valine to an isoleucine at amino acid residue 135 of the encoded protein. A third polymorphism is a change from a cytidine to a thymidine at position 41 of exon 8, referred to herein as "EX8C" where there is a cytidine at position 41 of exon 8 (the more common allele) or "EX8T," where there is a thymidine at position 41 of exon 8 (the less common allele). The change from a cytidine to a thymidine at position 41 of exon 8 does not result in a difference in the amino acid sequence of the encoded protein. A fourth polymorphism is a change from a cytidine to a thymidine at position 54 of intron 5, referred to herein as "IVS5C" where there is a cytidine at position 54 of intron 5 (the more common allele) or "IVS5T" where there is a thymidine at position 54 of intron 5 (the less common allele). A fifth polymorphism is a change from a cytidine to a guanine at position -41 of intron 10 (position -1 corresponds to the first nucleotide upstream of exon 11).

Specific allelic variants of these polymorphic regions are shown herein to be associated with specific disorders. In particular, the presence of a thymidine at position 41 in exon 8 (EX8T) was found to be associated with low plasma LDL levels in women and a

thymidine at position 54 of intron 5 (IVS5T) was found to be associated with a high BMI and high plasma LDL levels in women. In men, the presence of a thymidine at position 41 (EX8T) in exon 8, a thymidine at position 54 of intron 5 (IVS5T) and/or an adenine at position 146 of exon 1 was found to be associated with a high plasma HDL level. In addition, the presence of a cytidine at position 41 in exon 8 (EX8C) was found to be associated with low HDL levels in both women and men in three ethnically distinct populations. The presence of a thymidine at position 54 of intron 5 (IVS5T) was found to be associated with low HDL in women in three ethnically distinct populations. Furthermore, in women, the presence of both a cytidine at position 41 in exon 8 (EX8C) and a thymidine at position 54 of intron 5 (IVS5T) was found to be associated with a four-fold increase in the odds of having low HDL across three ethnically distinct populations. Since abnormal lipid, lipoprotein levels, and BMI may be associated with obesity, cachexia, diabetes, cardiovascular disease, gallstone formation and other disorders, SR-BI polymorphic variants are directly or indirectly associated with obesity, cachexia, diabetes, cardiovascular disease, gallstone formation and other disorders. SR-B1 variants, *e.g.*, EX8 and IVS5, may be used to predict the effect of hormone replacement therapy (HRT) on HDL levels in female subjects, *e.g.*, postmenopausal female subjects.

In one embodiment, the invention provides isolated nucleic acids comprising an intronic sequence from an SR-BI gene. In a preferred embodiment, the SR-BI gene is a human gene. In another preferred embodiment, the nucleic acid of the invention has a nucleotide sequence set forth in Figure 2A-G or in any of the intronic sequences in SEQ ID Nos. 1-121, complements thereof, or homologues thereof. In yet another embodiment, the intronic sequence of the nucleic acid is capable of hybridizing under an appropriate stringency to a nucleic acid having an intronic nucleotide sequence set forth in any of SEQ ID Nos. 1-121 or complements thereof.

Other preferred nucleic acids of the invention comprise at least an allelic variant of a polymorphic region. A preferred allele has a polymorphic region that is located in an exon and comprises, *e.g.*, a nucleotide sequence set forth in SEQ ID NO: 65, SEQ ID NO: 95, or SEQ ID NO:96 or a polymorphic region that is located in an intron and comprises, *e.g.*, a nucleotide sequence set forth in SEQ ID NO: 66 or SEQ ID NO: 97. The isolated nucleic acid preferably comprises from about 15 to about 30 nucleotides and can comprise, *e.g.*, a

nucleotide sequence selected from the group consisting of SEQ ID NO: 41 to SEQ ID NO: 64 and SEQ ID NO: 89 to SEQ ID NO: 94. The isolated nucleic acid can be double stranded or single stranded and can further comprise a label.

The nucleic acids of the invention can be used, *e.g.*, in prognostic, diagnostic, and therapeutic methods. For example, the nucleic acids of the invention can be used as probes or primers to determine whether a subject has or is at risk of developing a disease or disorder associated with a specific allelic variant of an SR-BI polymorphism, *e.g.*, a disease or disorder associated with an aberrant SR-BI activity, *e.g.*, obesity, diabetes, or cardiovascular disease.

The invention further describes vectors comprised of the claimed nucleic acids; host cells transfected with said vectors whether prokaryotic or eukaryotic; and transgenic non-human animals which contain a heterologous form of a functional or non-functional SR-BI allele described herein. Such a transgenic animal can serve as an animal model for studying, *e.g.*, the effect of specific allelic variations, including mutations of an SR-BI gene or for use in drug screening or recombinant protein production.

The invention further provides methods for determining the molecular structure of at least a portion of an SR-BI gene. In a preferred embodiment, the method comprises contacting a sample nucleic acid comprising an SR-BI gene sequence with a probe or primer having a sequence which is complementary to an SR-BI gene sequence and comparing the molecular structure of the sample nucleic acid with the molecular structure of a control (known) SR-BI gene (*e.g.*, an SR-BI gene from a human not afflicted with a cardiovascular condition or a disease associated with an aberrant SR-BI activity). The method of the invention can be used for example in determining the molecular structure of at least a portion of an exon, an intron, or the promoter. In a preferred embodiment, the method comprises determining the identity of at least one nucleotide. In even more preferred embodiments, the nucleotide is nucleotide 146 of exon 1, nucleotide 119 of exon 3, nucleotide 41 of exon 8, nucleotide 54 of intron 5, and/or nucleotide -41 of intron 10. In another preferred embodiment, the method comprises determining the nucleotide content of at least a portion of an SR-BI gene, such as by sequence analysis. In yet another embodiment, determining the molecular structure of at least a portion of an SR-BI gene is carried out by single-stranded conformation polymorphism. In yet another embodiment, the method is an oligonucleotide ligation assay. Other methods within the scope of the invention for determining the molecular

structure of at least a portion of an SR-BI gene include hybridization of allele-specific oligonucleotides, sequence specific amplification, and primer specific extension.

In at least some of the methods of the invention, the probe or primer is allele specific. Preferred probes or primers are single stranded nucleic acids, which optionally are labeled.

The methods of the invention can be used for determining the identity of the allelic variant of a polymorphic region of a human SR-BI gene present in a subject. For example, the method of the invention can be useful for determining whether a subject has, or is at risk of developing, a disease or condition associated with a specific allelic variant of a polymorphic region in the human SR-BI gene. In one embodiment, the disease or condition is characterized by an aberrant SR-BI activity, such as an aberrant SR-BI protein level, which can result from an aberrant expression of an SR-BI gene. The disease or condition can be an abnormal lipid metabolism, inappropriate lipid or lipoprotein level, an abnormal body mass index, atherosclerosis, or gallstone formation. Accordingly, the invention provides methods for predicting or diagnosing cardiovascular diseases or diabetes, and other diseases associated with an aberrant SR-BI activity.

For example, a female subject having the more common allele (a cytidine) at residue 41 of exon 8 of SR-BI (EX8C) has or is likely to have a tendency of having or developing higher LDL levels than a female subject having a thymidine at that position, thereby being at a higher risk of developing a cardiovascular disease. A female subject having the less common allele (a thymidine) at residue 54 of intron 5 (IVS5T) has or is likely to have or to develop a high BMI and/or high LDL levels, relative to a female subject having a cytidine at that position; and a male subject having the more common allele at residue 41 of exon 8 (EX8C), the more common allele at residue 54 of intron 5 (IVS5C), and the more common allele at residue 146 of exon 1 is likely to have or to develop lower HDL levels relative to a subject male having the less common alleles of at least one of these polymorphic regions, thereby being at a higher risk of developing a cardiovascular disease. Furthermore, a female or a male subject with the more common allele (cytidine) at residue 41 of exon 8 of SR-BI (EX8C) has or is likely to have a tendency of having or developing lower HDL levels than a female or a male subject having the less common allele (thymidine) at that position (EX8T), thereby being at a higher risk of developing a cardiovascular disease. A female subject with the less common allele (thymidine) at residue 54 of intron 5 (IVS5T) has or is likely to have

or develop low HDL levels, relative to a female subject having the more common allele (cytidine) at that position (IVS5C), thereby being at a higher risk of developing a cardiovascular disease; a female subject having both EX8C and IVS5T has greater than four-fold increased odds of having or developing low HDL levels as compared to a female subject having EX8T and IVS5C, thereby being at a higher risk of developing a cardiovascular disease.

The methods of the invention can also be used in selecting the appropriate drug to administer to a subject to treat a disease or condition, such as an abnormal lipid metabolism, inappropriate lipid level, a cardiovascular disease such as atherosclerosis, gallstone formation, diabetes, or an abnormal body mass index. In fact, specific allelic variants of SR-BI polymorphic regions may be associated with a specific response of an individual having such an allele to a specific drug. For example, a specific SR-BI allele may encode an SR-BI protein having a modified affinity for certain types of molecules, e.g, lipids. Accordingly, the action of a drug necessitating interaction with an SR-BI protein will be different in individuals carrying such an SR-BI allele.

In a further embodiment, the invention provides a method for treating a subject having a disease or condition associated with a specific allelic variant of a polymorphic region of an SR-BI gene. In one embodiment, the method comprises (a) determining the identity of the allelic variant; and (b) administering to the subject a compound that compensates for the effect of the specific allelic variant. In a preferred embodiment, the specific allelic variant is a mutation. The mutation can be located, *e.g.*, in a promoter region, an intron, or an exon of the gene. In one embodiment, the compound modulates (i.e., agonizes or antagonizes) SR-BI protein levels. In a preferred embodiment, the compound is selected from the group consisting of a nucleic acid, a protein, a peptidomimetic, or a small molecule. The compound can be, for example, an SR-BI protein. Thus, *e.g.*, if a female subject has the more common allele of residue 41 of exon 8 (EX8C), high LDL levels and resulting cardiovascular disorders can be treated, prevented from occurring or can be reduced, by administering to the subject a pharmaceutically effective amount of a compound which reduces LDL level to a normal LDL level. Similarly, if a female subject has the less common allele of residue 54 of intron 5 (IVS5T), a high BMI and/or LDL level and consequences thereof, such as diabetes and cardiovascular disorders, can be treated, prevented from occurring or can be reduced, by

administering to the subject a pharmaceutically effective amount of a compound which reduces the BMI and/or the LDL levels. If, on the other hand, a male subject has the more common allele at residue 41 of exon 8 (EX8C), the more common allele at residue 54 of intron 5 (IVS5C), and the more common allele at residue 146 of exon 1, development of low HDL levels can be treated, prevented, or increased by administering to the subject a pharmaceutically effective amount of a compound that increases HDL levels to normal levels. Likewise, if a female or a male subject has the more common allele at residue 41 of exon 8 (EX8C), development of low HDL levels and resulting cardiovascular disorders can be treated or prevented from occurring by administering to the subject a pharmaceutically effective amount of a compound which increases HDL level to normal HDL levels. Similarly, if a female subject has the less common allele of residue 54 of intron 5 (IVS5T), low HDL and resulting cardiovascular disorders can be treated or prevented from occurring, by administering to the subject a pharmaceutically effective amount of a compound which increases HDL level to a normal HDL level. Furthermore, if a female subject has both the more common allele at residue 41 of exon 8 (EX8C) and the less common allele of residue 54 of intron 5 (IVS5T), low HDL levels and resulting cardiovascular disorders can be treated or prevented from occurring, by administering to the subject a pharmaceutically effective amount of a compound which increases HDL level to a normal HDL level.

Since the effect of the presence of the variants are influenced by gender, the identification of one or more SR-B1 variants in a subject may be used to predict the effect of hormone replacement therapy (HRT). For example, if a female subject has one or more SR-B1 variants which have been associated with low HDL (*e.g.,* EX8C or IVS5T), but has normal HDL, HRT may cause low HDL in that subject. Likewise, if a female subject does not have either variant associated with low HDL (*e.g.,* EX8C or IVS5T), then it can be predicted that HRT would not cause low HDL in that subject. Therefore, in another embodiment, the invention provides a method of predicting the effect of HRT on a female subject with normal HDL levels, wherein the identification of allelic variants of the SR-B1 gene which are associated with abnormally low HDL levels in females results in a prediction that HRT will result in abnormally low HDL levels.

The invention also provides probes and primers comprising substantially purified oligonucleotides, which correspond to a region of nucleotide sequence which hybridizes to

at least 6 consecutive nucleotides of the sequence set forth as SEQ ID Nos: 1, 2, or 3 or to the complement of the sequences set forth as SEQ ID Nos: 1, 2, or 3; or naturally occurring mutants thereof. In preferred embodiments, the probe/primer further includes a label group attached thereto, which is capable of being detected.

In another embodiment, the invention provides a kit for amplifying and/or for determining the molecular structure of at least a portion of an SR-BI gene, comprising a probe or primer capable of hybridizing to an SR-BI gene and instructions for use. In one embodiment, the probe or primer is capable of hybridizing to an SR-BI intron. In another embodiment, the probe or primer is capable of hybridizing to an allelic variant of a polymorphic region. In a preferred embodiment, the polymorphic region is located in an exon, such as exon 1, 3, or 8 or in an intron, such as intron 5 or 10. In a preferred embodiment, determining the molecular structure of a region of an SR-BI gene comprises determining the identity of the allelic variant of the polymorphic region. Determining the molecular structure of at least a portion of an SR-BI gene can comprise determining the identity of at least one nucleotide or determining the nucleotide composition, *e.g.*, the nucleotide sequence.

A kit of the invention can be used, *e.g.*, for determining whether a subject has or is at risk of developing a disease associated with a specific allelic variant of a polymorphic region of an SR-BI gene, *e.g.*, EX8C or IVS5T. In a preferred embodiment, the invention provides a kit for determining whether a subject has or is at risk of developing a disease or condition associated with abnormal lipid metabolism, inappropriate lipid or lipoprotein levels, a cardiovascular disease such as atherosclerosis, gallstone formation, diabetes, or an abnormal body mass index. The disease or condition can be associated with an aberrant SR-BI activity, which can result, *e.g.*, from a mutation in the SR-BI gene. The kit of the invention can also be used in selecting the appropriate drug to administer to a subject to treat a disease or condition, such as a disease or condition set forth above. In fact, pharmacogenetic studies have shown that the genetic background of individuals play a role in determining the response of an individual to a specific drug. Thus, determining the allelic variants of SR-BI polymorphic regions of an individual can be useful in predicting how an individual will respond to a specific drug, *e.g.*, a drug for treating a disease or disorder associated with an

aberrant SR-BI activity and/or a cardiovascular disease or a disease associated with an aberrant lipid level.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

3. Brief Description of the Figures

Figure 1 is a schematic depiction of the chromosomal structure of the human SR-BI gene indicating the introns (1 through 12) and exons (I-XIII). Black boxes represent coding exons (exons I-XII) and the white box represents the non-coding exon (exon XIII) boxed and the nucleotides in the newly identified alleles are indicated.

Figure 2A-G represents the nucleotide sequence of the exons (underlined sequence) of the human SR-BI gene, portions of the introns which are adjacent to the exons, and 3' end of the promoter sequence (SEQ ID Nos. 5-40). The putative 5' end of the cDNA, as predicted by GRAIL is indicated in italics. The TATA-like box is indicated in italics and is boxed. Bold sequences correspond to the nucleotide sequence or the complement of the nucleotide sequence of preferred primers for amplifying each of the exons or a promoter region. The nucleotide polymorphisms in exons 1, 3, and 8 and introns 5 and 10 are boxed.

Figure 3A-B shows the nucleotide sequence of the full length human SR-BI cDNA (SEQ ID NO: 1) and the position of introns 1-12 relative to the nucleotide sequence of the exons. The nucleotide polymorphisms in exons 1, 3, and 8 are boxed.

Figure 4 is a graphic representing the mean LDL-C differences (+/- 95% Confidence intervals) between SR-BI genotypes carrying variant alleles and the wild-type genotype (111/111) in women. *significantly different from 111/111 ($p<0.030$). The differences between genotypes sharing letters are statistically significant (a: p=0.001; b: p=0.016; c: p=0.004).

Figure 5 is a graphic representing the mean LDL-C differences (+/- 95% Confidence intervals) between SR-BI genotypes carrying variant alleles and the wild-type genotype (111/111) in men. *significantly different from 111/111 ($p<0.030$). The differences between genotypes sharing letters are statistically significant (a: p=0.001; b: p=0.016; c: p=0.004).

Figure 6 is a graphic representing mean HDL differences (+/- 95% Confidence intervals) between SR-BI genotypes carrying variant alleles and the wild-type genotype (111/111) in women.

Figure 7 is a graphic representing mean HDL differences (+/- 95% Confidence intervals) between SR-BI genotypes carrying variant alleles and the wild-type genotype (111/111) in men. *significantly different from 111/111($p<0.040$).

Figure 8 is a graphic representing mean BMI differences (+/- 95% Confidence intervals) between SR-BI genotypes carrying variant alleles and the wild-type genotype (111/111) in women. *significantly different from 111/111 ($p=0.020$). The differences between genotypes sharing letters are statistically significant (a: $p=0.007$; b: $p=0.005$; c: $p=0.004$).

Figure 9 is a graphic representing mean BMI differences (+/- 95% Confidence intervals) between SR-BI genotypes carrying variant alleles and the wild-type genotype (111/111) in men.

4. Detailed Description of the Invention

4.1. General

The present invention is based at least in part on the discovery of the genomic structure of the human SR-BI gene and on the identification of polymorphic regions within the gene which correlate with specific diseases or conditions, such as an abnormal BMI or an abnormal lipoprotein (i.e., HDL or LDL) level. The present invention is based, at least in part, on the on the discovery of significant associations between two common polymorphisms, *e.g.*, position 41 of exon 8 (EX8) and position 54 of intron 5 (IVS5), within the SR-B1 gene and low HDL cholesterol (HDL) in three ethnically diverse populations. The three ethnically diverse populations consisted of two Scandanavian populations, one from Finland and one from southern Sweden, and a third population of Israelis of Ashkenazi Jewish origin. These subjects were drawn from collections of nuclear families ascertained from non-insulin dependent (type-2) diabetes mellitus (T2DM).

Both EX8C (*e.g.*, the more common allele at residue 41 of exon 8), and IVS5T (*e.g.*, the less common allele at residue 54 of intron 5) alone were positively associated with low HDL in female subjects in all three populations. In male subjects, the presence of EX8C

alone was shown to have a positive association with low HDL across all three populations. For male subjects, the presence of IVS5T alone was inversely associated with low HDL levels. The associations between SR-B1 EX8 and IVS5 alleles and low HDL levels are presented in Table I for women and Table II for men, set forth below.

TABLE I . Association between SR-B1 EX8 and IVS5 status and low HDL in women.

	EX8C carriers				IVS5T carriers			
Population	Cases	Controls	OR*	P value	Cases	Controls	OR	P value
Ashkenazi	79%	59%			18%	10%		
unadjusted			2.59	.003			2.06	.11
adjusted †			2.50	.008			2.42	.065
Finnish	88%	71%			10%	2%		
unadjusted			2.92	.013			4.95	.13
adjusted			3.69	.015			3.33	.28
Swedish	83%	68%			33%	13%		
unadjusted			2.33	.054			3.34	.013
adjusted			2.75	.044			2.73	.069
Combined	83%	65%			18%	9%		
unadjusted			2.66	<.00001			2.36	.006
adjusted			2.84	<.0001			2.32	.010

* Reference groups are homozygous TT for EX8 and homozygous CC for the IVS5 locus; OR= Odds ratio.

† All adjusted odds ratios control for T2DM, BMI, TG and age.

TABLE II. Association between SR-B1 EX8 and IVS5 status and low HDL in men.

	EX8C carriers				IVS5T carriers			
Population	Cases	Controls	OR*	P value	Cases	Controls	OR	P value
Ashkenazi	75%	61%			17%	17%		
unadjusted			1.95	.048			0.97	.95
adjusted †			2.11	.038			1.31	.55
Finnish	88%	85%			4%	10%		
unadjusted			1.38	.50			0.37	.16
adjusted			1.54	.42			0.43	.30
Swedish	84%	64%			8%	23%		
unadjusted			2.82	.014			0.30	.024
adjusted			3.02	.020			0.25	.020
Combined	81%	71%			11%	16%		
unadjusted			1.81	.008			0.62	.084
adjusted			1.79	.015			0.64	.15

*Reference groups are homozygous TT for EX8 and homozygous CC for the IVS5 locus; OR= Odds ratio.
† All adjusted odds ratios control for T2DM, BMI, TG and age.

Among female subjects, the presence of both EX8C and IVS5T results in four-fold increased odds of having or developing low HDL as compared to those without either variant. Male carriers of both EX8C and IVS5T, were no more likely to have low HDL than individuals without either variant. The combined effect of both variants on HDL level is shown in Table III, below.

Table III. Effect of both EX8 and IVS5 status on low HDL in female and male subjects from the three populations combined.

WOMEN						
EX8C	IVS5T	cases	controls	OR	95% C.I.	P value
+	+	53	15	4.79	(2.30, 10.07)	<.00001
+	-	189	105	2.44	(1.53, 3.89)	<.0001
-	-/+ *	48	65	1.00	-	-
			chi square $p<2x10^{-6}$			
MEN						
EX8C	IVS5T	cases	controls	OR	95% C.I.	P value
+	+	28	29	1.08	(0.54, 2.18)	.95
+	-	184	106	1.95	(1.21, 3.14)	<.01
-	-/+ †	49	55	1.00	-	-
			chi square p<.006			

* includes only one person who is +
† includes only two people who are +

As described above, female carriers of both EX8C and IVS5T had the highest odds of having low HDL. However, even carriers of EX8C who lacked the IVS5T variant were at increased odds of having low HDL. Considering the nature of these variants (EX8 is silent and IVS5 intronic), the results suggest that neither EX8C nor IVS5T are the causative polymorphism underlying the association with low HDL. The underlying variant(s) may be found within haplotypes represented by different combinations of alleles at both EX8 and IVS5. As used herein, a "haplotype" is a set or pattern of polymorphisms which are conserved and which confer a certain phenotype. Two different risk patterns are evident in women. The first pattern, conferring the greater than four-fold increased odds of low HDL, is defined by presence of both EX8C and IVS5T. The second pattern, conferring increased (>2.4-fold) odds of low HDL, is defined by presence of EX8C and absence of IVS5T. In men, only the second pattern was associated with low HDL, conferring about a two-fold increased odds. This association is consistent with what was found in women. However, unlike women, men with the first pattern were no more likely to have low HDL than those who lacked both variant alleles.

It is well known that HDL levels are affected by sex hormone status. Furthermore, the expression of SR-B1 is known to be regulated by estrogen. Estrogen treatment of rats

has been shown to downregulate SR-B1 in the liver (Fluiter K, et al. (1998) *J Biol Chem.* 273:8434-8). Moreover, overexpression of SR-B1 in the liver has been demonstrated to result in a pronounced fall in plasma HDL (Kozarsky KF (1997) *Nature* 387:414-7). It is possible that the downregulation of SR-B1 by estrogen is impaired by a genetic variant in SR-B1, resulting in an increased expression of SR-B1 and therefore lower HDL levels in women. This same effect may not be apparent in men as estrogen does not play as key a role in the modulation of HDL.

The possible interaction between the SR-B1 variants and hormonal status has implications for the treatment of females with hormone replacement therapy (HRT). It is possible that SR-B1 variants may modulate the effect of HRT on HDL levels in women. For example, a postmenopausal woman may have the EX8C and IVS5T variants, but may have normal HDL. However, treatment with HRT may cause low HDL. Therefore, in females, the identification of SR-B1 variants (*e.g.,* EX8C and/or IVS5T) may be used to predict the effect HRT would have on HDL level (*e.g.,* lowering HDL level). SR-B1 genotype may also have utility as a pharmacogenomic marker of response to lipid lowering therapies.

In women, the associations between SR-B1 variants and low HDL were consistently demonstrated across three different ethnic populations. In men, some population-specific effects were observed. The ability to reproduce an association in multiple populations adds to the validity of these results. It is unlikely that the observed associations were spurious due to population stratification in all three of these relatively homogenous groups. The replication of these associations in three distinct ethnic groups also suggests that the haplotypes carrying the underlying variants may be the same across populations.

The strongest association with HDL level is not with either variant alone, but rather the combination of genotypes at EX8 and IVS5. Therefore, the location of additional genetic variants in or around the SR-B1 locus, in linkage disequilibrium with IVS5 and EX8, may influence these associations. These additional genetic variants may include the variants described herein or other variants present within the SR-B1 gene.

The association of SR-B1 polymorphisms with low HDL remained significant after controlling for other covariates known to be associated with HDL including T2DM status,

BMI, age and triglyceride (TG) levels. This suggests that SR-B1 is an independent predictor of HDL levels. Most of the individuals were affected or related to individuals affected with T2DM. Combined low HDL and high TG are the hallmark dyslipidemia associated with T2DM. This combined phenotype may be a reasonable marker of insulin resistance (Jeppesen, et al (1997) *J. Arterioscler Thromb Vasc Biol.* 17:1114-20).

SR-B1 genetic variants showed strong associations with low HDL. The high frequency of the variants combined with their strong effect suggests that SR-B1 is a major genetic determinant of low HDL. The fraction of low HDL attributable to the EX8 variant was estimated to be 35% in women and 23% in men in the three populations tested. The association of SR-B1 variants with low HDL described herein is not only strong but highly significant and reproducible. Therefore, SR-B1 may be used as a genetic marker of plasma HDL cholesterol levels.

As shown in Figure 1, the human SR-BI gene is at least 50 kilobase pairs long and has 12 coding exons, one non-coding exon (exon 13), and 12 introns. The exons are numbered 1 to 13 from 5' to 3' and the introns are labeled 1 through 12 from 5' to 3'. Exon 1 corresponds to the first exon located downstream of the promoter and contains the initiation codon. Intron 1 is located immediately downstream of exon I (see Figure 1). The position of the introns relative to the nucleotide sequence of the full length cDNA encoding SR-BI is shown in Figure 2A-G. The nucleotide sequence of the human SR-BI cDNA, shown in Figure 3 and in SEQ ID NO: 1 encodes a protein of 509 amino acids. SEQ ID NO. 1 contains the nucleotide sequence of the cDNA disclosed in Calvo and Vega (1993) J. Biol. Chem. 268:18929, and contains in addition a complete 5' end. The amino acid sequence of the protein set forth in SEQ ID NO: 2 is identical to the Cla-I protein disclosed in Calvo and Vega (1993) J. Biol. Chem. 268:18929. As set forth in Calvo and Vega, *supra*, differential splicing of the human SR-BI gene also results in a short mRNA lacking 300 nucleotides located 126 nucleotides downstream of the initiation codon, i.e., lacking exons 2 and 3 set forth in Figure 3, which encodes a protein of 409 amino acids. The shorter protein is referred to herein as "splice variant". The nucleotide sequence of a full length cDNA encoding the splice variant is set forth in SEQ ID NO: 3 and the amino acid sequence of the SR-BI splice variant protein encoded by this nucleotide sequence is set forth in SEQ ID NO: 4. The splice variant is rare relative to the 509 amino acid SR-BI protein.

Figure 2A-G shows the nucleotide sequence of the 3' end of the SR-BI promoter. Additional promoter sequence is disclosed in U.S. Patent Application 08/812,204 by Acton, incorporated herein by reference.

Set forth below in Table IV are the locations and sizes of the exons in the human SR-BI gene relative to the nucleotide sequence of a full length cDNA encoding human SR-BI protein (SEQ ID NO: 1), in which nucleotide 1 corresponds to the first nucleotide in the isolated transcript. Table IV also indicates the portions of the human SR-BI protein encoded by each of these exons. Amino acid 1 is the initiating methionine. Also indicated is the length of the intron located downstream of each of the exons.

TABLE IV

	cDNA Nucleotide position	Amino acid position	Size of intron
Exon 1	1-244	1-42	intron 1: >2827
Exon 2	245-402	43-95	intron 2: 2429
Exon 3	403-544	95-142	intron 3: 567
Exon 4	545-748	143-210	intron 4: 2229
Exon 5	749-844	211-242	intron 5: 1580
Exon 6	845-960	243-281	intron 6: >10532
Exon 7	961-1127	281-337	intron 7: >3985
Exon 8	1228-1246	337-376	intron 8: >11321
Exon 9	1247-1320	377-401	intron 9: 7562
Exon 10	1321-1372	401-418	intron 10: 902
Exon 11	1373-1519	419-467	intron 11: 3547
Exon 12	1520-1648	468-509	intron 12: >4578
Exon 13	1649-2630		

Figure 2A-G shows the nucleotide sequence of portions of the introns which are adjacent to the exons. The nucleotide sequence of each of the exons and adjacent portions of introns shown in Figure 2A-G are set forth in SEQ ID Nos. 5 to 16. The portions of each of the introns shown in Figure 2A-G are set forth in SEQ ID Nos. 18 to 40. For convenience,

the identity of the sequences referred to as SEQ ID Nos. 1 to 40 are set forth below in Table V:

TABLE V

SEQ ID NO: 1	full length cDNA encoding human SR-BI;
SEQ ID NO: 2	full length amino acid sequence of human SR-BI protein;
SEQ ID NO: 3	full length cDNA encoding splice variant of human SR-BI (Calvo and Vega, *supra*);
SEQ ID NO: 4	full length amino acid sequence of splice variant of human SR-BI protein (Calvo and Vega, *supra*);
SEQ ID NO: 5	3' end of promoter, exon 1, and 5' end of intron 1;
SEQ ID NO: 6	3' end of intron 1, exon 2, and 5' end of intron 2;
SEQ ID NO: 7	3' end of intron 2, exon 3, and 5' end of intron 3;
SEQ ID NO: 8	3' end of intron 3, exon 4, and 5' end of intron 4;
SEQ ID NO: 9	3' end of intron 4, exon 5, and 5' end of intron 5;
SEQ ID NO: 10	3' end of intron 5, exon 6, and 5' end of intron 6;
SEQ ID NO: 11	3' end of intron 6, exon 7, and 5' end of intron 7;
SEQ ID NO: 12	3' end of intron 7, exon 8, and 5' end of intron 8;
SEQ ID NO: 13	3' end of intron 8, exon 9, and 5' end of intron 9;
SEQ ID NO: 14	3' end of intron 9, exon 10, and 5' end of intron 10;
SEQ ID NO: 15	3' end of intron 10, exon 11, and 5' end of intron 11;
SEQ ID NO: 16	3' end of intron 11, exon 12, and 5' end of intron 12;
SEQ ID NO: 17	3' end of promoter;
SEQ ID NO: 18	5' end of intron 1;
SEQ ID NO: 19	3' end of intron 1;
SEQ ID NO: 20	5' end of intron 2;
SEQ ID NO: 21	3' end of intron 2;
SEQ ID NO: 22	5' end of intron 3;
SEQ ID NO: 23	3' end of intron 3;
SEQ ID NO: 24	5' end of intron 4;
SEQ ID NO: 25	3' end of intron 4;

SEQ ID NO: 26	5' end of intron 5;
SEQ ID NO: 27	3' end of intron 5;
SEQ ID NO: 28	5' end of intron 6;
SEQ ID NO: 29	3' end of intron 6;
SEQ ID NO: 30	5' end of intron 7;
SEQ ID NO: 31	3' end of intron 7;
SEQ ID NO: 32	5' end of intron 8;
SEQ ID NO: 33	3' end of intron 8;
SEQ ID NO: 34	5' end of intron 9;
SEQ ID NO: 35	3' end of intron 9;
SEQ ID NO: 36	5' end of intron 10;
SEQ ID NO: 37	3' end of intron 10;
SEQ ID NO: 38	5' end of intron 11;
SEQ ID NO: 39	3' end of intron 11; and
SEQ ID NO: 40	5' end of intron 12.

An analysis of the human SR-BI gene in a population of individuals chosen because these individuals had a known age, known HDL and LDL levels, known body mass index, and known triglycerides and total cholesterol levels revealed the existence of several polymorphisms in the SR-BI gene in this population. These polymorphisms were identified by performing single stranded conformation polymorphism (SSCP) analysis of genomic DNA from independent individuals as described in Example 3 and in Example 6, using PCR primers complementary to intronic or promoter sequences surrounding each of the exons. The nucleotide sequence of these PCR primers (having SEQ ID Nos. 41-64) is shown in Table VII (in the Examples).

The results indicated the presence of at least five polymorphic regions in the human SR-BI gene in the population studied. The location and identity of these polymorphisms is indicated in Table VI.

TABLE VI Locations of polymorphisms in the human SR-BI gene.

poly-morphism	cDNA position	amino acid position	change	location
exon 1	4	2	Gly -> Ser	1 2 3 ATG (G/A)GC TGC
exon 3	403	135	Val -> Ile	135 136 137 (G/A)TC ATG CCC
intron 5	na	na	na	240 241 242 CTG AGC AAG gtgaggggcgagaggcgagggccctgt cgccagggagggggagggtgggcc(c/t)g (SEQ ID NO.:87)
exon 8	1050	350	none	350 351 352 (C/T)GA CCC GGT
intron 10	na	na	na	419 c(c/g)tgcggccccagctcatgtgtttgtcattctgtctcctcag AGC 420 421 GGG GCC (SEQ ID NO.:88)

The intron is defined as being after its corresponding exon (intron 1 is 3' of exon 1), cDNA position 1= the first base of the initiator methionine, the numbers above the sequences refer to the amino acid number, na = not applicable, lower case indicates intronic sequence, and the polymorphisms are in parentheses.

As can be seen in Table VI, one polymorphism is a change from a guanine to an adenine at position 146 in exon 1, which results in a change from a glycine to a serine at amino acid residue 2 of the encoded protein. The nucleotide sequence of this allele is set forth in SEQ ID NO: 95 (which is identical to SEQ ID NO: 5, except for nucleotide 146 of exon 1 which is an adenine). A second polymorphism is a change from a guanine to an adenine at position 119 in exon 3, which results in a change from a valine to an isoleucine at amino acid residue 135 of the encoded protein. The nucleotide sequence of this allele is set forth in SEQ ID NO: 96 (which is identical to SEQ ID NO: 7, except for nucleotide 119 of exon 3 which is an adenine).

A third polymorphism is a change from a cytidine to a thymidine at position 41 of exon 8, which does not result in a difference in the amino acid sequence of the encoded protein. The nucleotide sequence of exon 8 of this allele is set forth in SEQ ID NO: 65 (SEQ ID NO: 65 is identical to SEQ ID NO: 12, except for nucleotide 41 of the exon sequence which is a thymidine). About 35 % of the individuals of a Spanish population of 142 individuals were found to be homozygous for the allele having a cytidine at position 41 (i.e., SR-BI sequence originally disclosed); about 17% of the individuals were found to be homozygous for the allele

having a thymidine at position 41 of exon 8; and about 48% of the individuals were found to be heterozygous, i.e., having one allele having a cytidine at position 41 and one allele having a thymidine at position 41.

A fourth polymorphism is a change from a cytidine to a thymidine at position 54 of intron 5 (position 1 being defined as the first nucleotide in the intron). This nucleotide substitution destroys the ApaI restriction site which is present when the nucleotide at position 54 is a cytidine. The nucleotide sequence of the 5' end of intron 5 of this allele is set forth in SEQ ID NO: 66 (SEQ ID NO: 66 is identical to SEQ ID NO: 26, except for nucleotide 54 which is a thymidine).

A fifth polymorphism in the SR-BI gene is a change from a cytidine to a guanine at position -41 of intron 10 (position -1 corresponds to the first nucleotide upstream of exon 11). The nucleotide sequence of the 3' end of intron 10 of this allele is set forth in SEQ ID NO: 97 (SEQ ID NO: 97 is identical to SEQ ID NO: 15, except for nucleotide -41 of intron 10 which is a guanine).

Further analysis of the human SR-BI gene is likely to reveal the existence of yet other polymorphic regions. Such analysis can be performed using the methods described herein and genomic DNA from random subjects or subjects of families associated with specific diseases. For example, the polymorphism studies described herein can also be applied to populations in which cholesterol gallstones are prevalent. Accordingly, the invention provides materials and methods, such as nucleic acids (e.g, intronic sequences useful as probes or primers) for determining the identity of other allelic variants of an SR-BI polymorphic region. The invention also provides methods for determining the identity of the alleles of a specific polymorphic region of an SR-BI gene (see *e.g.*, Examples 1-4). Such methods can be used, for example, to determine whether a subject has or is at risk of developing a disease or condition associated with one or more specific alleles of polymorphic regions of an SR-BI gene (see *e.g.*, Example 5). In a preferred embodiment, the disease or condition is caused or contributed to by an aberrant SR-BI bioactivity. Other aspects of the invention are described below or will be apparent to one of skill in the art in light of the present disclosure.

4.2 Definitions

For convenience, the meaning of certain terms and phrases employed in the specification, examples, and appended claims are provided below.

The term "allele", which is used interchangeably herein with "allelic variant" refers to alternative forms of a gene or portions thereof. Alleles occupy the same locus or position on homologous chromosomes. When a subject has two identical alleles of a gene, the subject is said to be homozygous for the gene or allele. When a subject has two different alleles of a gene, the subject is said to be heterozygous for the gene. Alleles of a specific gene can differ from each other in a single nucleotide, or several nucleotides, and can include substitutions, deletions, and insertions of nucleotides. An allele of a gene can also be a form of a gene containing a mutation.

The term "allelic variant of a polymorphic region of an SR-BI gene" refers to a region of an SR-BI gene having one of several nucleotide sequences found in that region of the gene in other individuals.

Biological activity" or "bioactivity" or "activity" or "biological function", which are used interchangeably, for the purposes herein when applied to SR-BI means an effector or antigenic function that is directly or indirectly performed by an SR-BI polypeptide (whether in its native or denatured conformation), or by any subsequence (fragment) thereof. Biological activities include binding to a ligand, *e.g.*, a lipid or lipoprotein, such as LDL or modified forms thereof, or HDL or modified forms thereof. Other molecules which can bind an SR-BI receptor include anionic molecules, such as anionic phospholipids, negatively charged liposomes, and apoptotic cells. Another biological activity of an SR-BI protein includes endocytosis of a ligand interacting with the receptor. A biological activity is also intended to include binding to a protein, such as binding to the cytoplasmic domain of SR-BI. Yet other biological activities include signal transduction from the receptor, modulation of expression of genes responsive to binding of a ligand to an SR-BI receptor, and other biological activities, whether presently known or inherent. An SR-BI bioactivity can be modulated by directly affecting an SR-BI protein. Alternatively, an SR-BI bioactivity can be modulated by modulating the level of an SR-BI protein, such as by modulating expression of an SR-BI gene. Antigenic functions include possession of an epitope or antigenic site that is

capable of cross-reacting with antibodies raised against a naturally occurring or denatured SR-BI polypeptide or fragment thereof.

Biologically active SR-BI polypeptides include polypeptides having both an effector and antigenic function, or only one of such functions. SR-BI polypeptides include antagonist polypeptides and native SR-BI polypeptides, provided that such antagonists include an epitope of a native SR-BI polypeptide. An effector function of SR-BI polypeptide can be the ability to bind to a ligand, *e.g.*, a lipid or modified form thereof.

As used herein the term "bioactive fragment of a SR-BI protein" refers to a fragment of a full-length SR-BI protein, wherein the fragment specifically mimics or antagonizes the activity of a wild-type SR-BI protein. The bioactive fragment preferably is a fragment capable of binding to a second molecule, such as a ligand.

The term "an aberrant activity" or "abnormal activity", as applied to an activity of a protein such as SR-BI, refers to an activity which differs from the activity of the wild-type or native protein or which differs from the activity of the protein in a healthy subject, *e.g.*, a subject not afflicted with a disease associated with a specific allelic variant of an SR-BI polymorphism. An activity of a protein can be aberrant because it is stronger than the activity of its native counterpart. Alternatively, an activity can be aberrant because it is weaker or absent related to the activity of its native counterpart. An aberrant activity can also be a change in an activity. For example an aberrant protein can interact with a different protein relative to its native counterpart. A cell can have an aberrant SR-BI activity due to overexpression or underexpression of the gene encoding SR-BI. An aberrant SR-BI activity can result, *e.g.*, from a mutation in the gene, which results, *e.g.*, in lower or higher binding affinity of a lipid to the SR-BI protein encoded by the mutated gene. An aberrant SR-BI activity can also result from a lower or higher level of SR-BI receptor on cells, which can result, *e.g.*, from a mutation in the 5' flanking region of the SR-BI gene or any other regulatory element of the SR-BI gene, such as a regulatory element located in an intron. Accordingly, an aberrant SR-BI activity can result from an abnormal SR-BI promoter activity.

The terms "abnormal SR-BI promoter activity", "aberrant SR-BI promoter activity", "abnormal SR-BI transcriptional activity" and "aberrant SR-BI transcriptional activity", which are used interchangeably herein, refer to the transcriptional activity of an SR-BI promoter which differs from the transcriptional activity of the same promoter in a healthy

subject. An abnormal SR-BI activity can result from a higher or lower transcriptional activity than that in a healthy subject. An aberrant SR-BI promoter activity can result, *e.g.*, from the presence of a genetic lesion in a regulatory element, such as in a regulatory element located in the promoter. An "aberrant SR-BI promoter activity" is also intended to refer to the transcriptional activity of an SR-BI promoter which is functional (capable of inducing transcription of a gene to which it is operably linked) in tissues or cells in which the "natural" or wild-type SR-BI promoter is not functional or which is non functional in tissues or cells in which the "natural" or wild-type SR-BI promoter is functional. Thus, a tissue distribution of SR-BI in a subject which differs from the tissue distribution of SR-BI in a "normal" or "healthy" subject, can be the result of an abnormal transcriptional activity from the SR-BI promoter region. Such an abnormal transcriptional activity can result, *e.g.*, from one or more mutations in a promoter region, such as in a regulatory element thereof. An abnormal transcriptional activity can also result from a mutation in a transcription factor involved in the control of SR-BI gene expression.

The term "body mass index" or "BMI" refers to the ratio of weight (kg)/height $(m)^2$ and can be used to define whether a subject is overweight. Typically, a subject is underweight if he has a BMI<20; normal if he has a BMI of 20-25, overweight if he has a BMI of 25-30, obese if he has a BMI of 30-40 and severely obese if he has a BMI>40.

As used herein, a subject has an "abnormal body mass" or "abnormal body mass index" or "aberrant body mass" or "aberrant body mass index" if his body mass index is outside the range defined for a healthy or normal subject, i.e., BMI of 20-25. A disorder of body mass include any disorder affecting the body mass of a subject such that his body mass is outside the normal range. For example, obesity is a disorder of body mass. Wasting is also a disorder of body mass. An abnormal body mass index can have a hormonal origin, *e.g.*, in premenopausal women.

The term "cardiovascular disorder" refers to a disease or disorder of the cardiovascular system and includes ischemia, restenosis, congestive heart failure, and atherosclerosis.

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding

generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

As used herein, the term "gene" or "recombinant gene" refers to a nucleic acid molecule comprising an open reading frame and including at least one exon and (optionally) an intron sequence. The term "intron" refers to a DNA sequence present in a given gene which is spliced out during mRNA maturation.

"Hormone replacement therapy" (HRT) refers to the administration of female hormones to a female subject, *e.g.*, a postmenopausal female subject.

"Homology" or "identity" or "similarity" refers to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences. An "unrelated" or "non-homologous" sequence shares less than 40 % identity, though preferably less than 25 % identity, with one of the sequences of the present invention.

The term " a homologue of a nucleic acid" refers to a nucleic acid having a nucleotide sequence having a certain degree of homology with the nucleotide sequence of the nucleic acid or complement thereof. A homologue of a double stranded nucleic acid having SEQ ID NO: x is intended to include nucleic acids having a nucleotide sequence which has a certain degree of homology with SEQ ID NO: x or with the complement thereof. Preferred homologues of nucleic acids are capable of hybridizing to the nucleic acid or complement thereof.

The term "interact" as used herein is meant to include detectable interactions between molecules, such as can be detected using, for example, a hybridization assay. The term interact is also meant to include "binding" interactions between molecules. Interactions may be, for example, protein-protein, protein-nucleic acid, protein-small molecule or small molecule-nucleic acid in nature.

The term "intronic sequence" or "intronic nucleotide sequence" refers to the nucleotide sequence of an intron or portion thereof .

The term "isolated" as used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs or RNAs, respectively, that are present in the natural source of the macromolecule. The term isolated as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and would not be found in the natural state. The term "isolated" is also used herein to refer to polypeptides which are isolated from other cellular proteins and is meant to encompass both purified and recombinant polypeptides.

The term "lipid" shall refer to a fat or fat-like substance that is insoluble in polar solvents such as water. The term "lipid" is intended to include true fats (*e.g.* esters of fatty acids and glycerol); lipids (phospholipids, cerebrosides, waxes); sterols (cholesterol, ergosterol) and lipoproteins (*e.g.* HDL, LDL and VLDL).

The term "locus" refers to a specific position in a chromosome. For example, a locus of an SR-BI gene refers to the chromosomal position of the SR-BI gene.

The term "modulation" as used herein refers to both upregulation, (i.e., activation or stimulation), for example by agonizing; and downregulation (i.e. inhibition or suppression), for example by antagonizing of a bioactivity (*e.g.* expression of a gene).

The term "molecular structure" of a gene or a portion thereof refers to the structure as defined by the nucleotide content (including deletions, substitutions, additions of one or more nucleotides), the nucleotide sequence, the state of methylation, and/or any other modification of the gene or portion thereof.

The term "mutated gene" refers to an allelic form of a gene, which is capable of altering the phenotype of a subject having the mutated gene relative to a subject which does not have the mutated gene. If a subject must be homozygous for this mutation to have an altered phenotype, the mutation is said to be recessive. If one copy of the mutated gene is sufficient to alter the genotype of the subject, the mutation is said to be dominant. If a subject has one copy of the mutated gene and has a phenotype that is intermediate between that of a homozygous and that of a heterozygous (for that gene) subject, the mutation is said to be co-dominant.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, derivatives, variants and analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides. Deoxyribonucleotides include deoxyadenosine, deoxycytidine, deoxyguanosine, and deoxythymidine. For purposes of clarity, when referring herein to a nucleotide of a nucleic acid, which can be DNA or an RNA, the terms "adenosine", "cytidine", "guanosine", and thymidine" are used. It is understood that if the nucleic acid is RNA, a nucleotide having a uracil base is uridine.

The term "nucleotide sequence complementary to the nucleotide sequence set forth in SEQ ID NO: x" refers to the nucleotide sequence of the complementary strand of a nucleic acid strand having SEQ ID NO: x. The term "complementary strand" is used herein interchangeably with the term "complement". The complement of a nucleic acid strand can be the complement of a coding strand or the complement of a non-coding strand. When referring to double stranded nucleic acids, the complement of a nucleic acid having SEQ ID NO: x refers to the complementary strand of the strand having SEQ ID NO: x or to any nucleic acid having the nucleotide sequence of the complementary strand of SEQ ID NO: x. When referring to a single stranded nucleic acid having the nucleotide sequence SEQ ID NO: x, the complement of this nucleic acid is a nucleic acid having a nucleotide sequence which is complementary to that of SEQ ID NO: x. The nucleotide sequences and complementary sequences thereof are always given in the 5' to 3' direction. The term "complement" and "reverse complement" are used interchangeably herein.

A "non-human animal" of the invention can include mammals such as rodents, non-human primates, sheep, goats, horses, dogs, cows, chickens, amphibians, reptiles, etc. Preferred non-human animals are selected from the rodent family including rat and mouse, most preferably mouse, though transgenic amphibians, such as members of the *Xenopus* genus, and transgenic chickens can also provide important tools for understanding and identifying agents which can affect, for example, embryogenesis and tissue formation. The term "chimeric animal" is used herein to refer to animals in which an exogenous sequence is found, or in which an exogenous sequence is expressed in some but not all cells

of the animal. The term "tissue-specific chimeric animal" indicates that an exogenous sequence is present and/or expressed or disrupted in some tissues, but not others.

The term "operably linked" is intended to mean that the promoter is associated with the nucleic acid in such a manner as to facilitate transcription of the nucleic acid from the promoter.

The term "polymorphism" refers to the coexistence of more than one form of a gene or portion thereof. A portion of a gene of which there are at least two different forms, i.e., two different nucleotide sequences, is referred to as a "polymorphic region of a gene". A polymorphic region can be a single nucleotide, the identity of which differs in different alleles. A polymorphic region can also be several nucleotides long.

A "polymorphic gene" refers to a gene having at least one polymorphic region.

The terms "protein", "polypeptide" and "peptide" are used interchangeably herein when referring to a gene product.

The term "recombinant protein" refers to a polypeptide which is produced by recombinant DNA techniques, wherein generally, DNA encoding the polypeptide is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein.

A "regulatory element", also termed herein "regulatory sequence is intended to include elements which are capable of modulating transcription from a basic promoter and include elements such as enhancers and silencers. The term "enhancer", also referred to herein as "enhancer element", is intended to include regulatory elements capable of increasing, stimulating, or enhancing transcription from a basic promoter. The term "silencer", also referred to herein as "silencer element" is intended to include regulatory elements capable of decreasing, inhibiting, or repressing transcription from a basic promoter. Regulatory elements are typically present in 5' flanking regions of genes. However, regulatory elements have also been shown to be present in other regions of a gene, in particular in introns. Thus, it is possible that SR-BI genes have regulatory elements located in introns, exons, coding regions, and 3' flanking sequences. Such regulatory elements are also intended to be encompassed by the present invention and can be identified by any of the assays that can be used to identify regulatory elements in 5' flanking regions of genes.

The term "regulatory element" further encompasses "tissue specific" regulatory elements, i.e., regulatory elements which effect expression of the selected DNA sequence preferentially in specific cells (*e.g.*, cells of a specific tissue). Gene expression occurs preferentially in a specific cell if expression in this cell type is significantly higher than expression in other cell types. The term "regulatory element" also encompasses non-tissue specific regulatory elements, i.e., regulatory elements which are active in most cell types. Furthermore, a regulatory element can be a constitutive regulatory element, i.e., a regulatory element which constitutively regulates transcription, as opposed to a regulatory element which is inducible, i.e., a regulatory element which is active primarily in response to a stimulus. A stimulus can be, *e.g.*, a molecule, such as a hormone, cytokine, heavy metal, phorbol ester, cyclic AMP (cAMP), or retinoic acid.

Regulatory elements are typically bound by proteins, *e.g.*, transcription factors. The term "transcription factor" is intended to include proteins or modified forms thereof, which interact preferentially with specific nucleic acid sequences, i.e., regulatory elements, and which in appropriate conditions stimulate or repress transcription. Some transcription factors are active when they are in the form of a monomer. Alternatively, other transcription factors are active in the form of a dimer consisting of two identical proteins or different proteins (heterodimer). Modified forms of transcription factors are intended to refer to transcription factors having a postranslational modification, such as the attachment of a phosphate group. The activity of a transcription factor is frequently modulated by a postranslational modification. For example, certain transcription factors are active only if they are phosphorylated on specific residues. Alternatively, transcription factors can be active in the absence of phosphorylated residues and become inactivated by phosphorylation. A list of known transcription factors and their DNA binding site can be found, *e.g.*, in public databases, *e.g.*, TFMATRIX Transcription Factor Binding Site Profile database.

As used herein, the term "specifically hybridizes" or "specifically detects" refers to the ability of a nucleic acid molecule of the invention to hybridize to at least approximately 6, 12, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130 or 140 consecutive nucleotides of either strand of an SR-BI gene.

"SR-BI" or "SR-BI receptor" refers to a class B scavenger receptor that has been shown to bind HDL cholesterol and mediate uptake into cells (Acton, S. et al., *Science*

271:518-520). SR-BI has also been shown to bind with high affinity to modified proteins (*e.g.* acetylated LDL, oxidized LDL, maleylated bovine serum albumin) and native LDL (Acton, et al., (1994) *J. Biochem. 269*:21003-21009). Further, SR-BI has been shown to bind anionic phospholipids, such as phosphatidylserine and phosphatidylinositol, but not zwitterionic phospholipids, such as phosphatidylcholine, phosphatidylethanolamine and sphingomyelin. Competition studies suggest that anionic phospholipids bind to SR-BI at a site close to or identical with the sites of native and modified LDL binding and that the interaction may involve polyvalent binding via multiple anionic phospholipid molecules (Rigotti, A.., S. Acton and M. Krieger (1995) *J. Biochem 270*:16221-16224). SR-BI has also been shown to bind to negatively charged liposomes and apoptotic cells. The human SR-BI protein is described in Calvo et al. (1993) J. Biol. Chem. 268:18929 and hamster SR-BI is described in International Patent Application Number WO 96/00288 entitled "Class B1 and C1 Scavenger Receptors" by Acton, S. et al.

The term "SR-BI therapeutic" refers to various forms of SR-BI polypeptides, as well as peptidomimetics, nucleic acids, or small molecules, which can modulate at least one activity of an SR-BI by mimicking or potentiating (agonizing) or inhibiting (antagonizing) the effects of a naturally-occurring SR-BI polypeptide. An SR-BI therapeutic which mimics or potentiates the activity of a wild-type SR-BI polypeptide is a "SR-BI agonist". Conversely, an SR-BI therapeutic which inhibits the activity of a wild-type SR-BI polypeptide is a "SR-BI antagonist". SR-BI therapeutics can be used to treat diseases which are associated with a specific SR-BI allele which encodes a protein having an amino acid sequence that differs from that of the wild-type SR-BI protein.

As used herein, the term "transfection" means the introduction of a nucleic acid, *e.g.*, an expression vector, into a recipient cell by nucleic acid-mediated gene transfer. The term "transduction" is generally used herein when the transfection with a nucleic acid is by viral delivery of the nucleic acid. "Transformation", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA, and, for example, the transformed cell expresses a recombinant form of a polypeptide or, in the case of anti-sense expression from the transferred gene, the expression of a naturally-occurring form of the recombinant protein is disrupted.

As used herein, the term "transgene" refers to a nucleic acid sequence which has been introduced into a cell. Daughter cells deriving from a cell in which a transgene has been introduced are also said to contain the transgene (unless it has been deleted). A transgene can encode, *e.g.*, a polypeptide, or an antisense transcript, partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (*e.g.*, it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). Alternatively, a transgene can also be present in an episome. A transgene can include one or more transcriptional regulatory sequence and any other nucleic acid, (*e.g.* intron), that may be necessary for optimal expression of a selected nucleic acid.

A "transgenic animal" refers to any animal, preferably a non-human animal, *e.g.* a mammal, bird or an amphibian, in which one or more of the cells of the animal contain heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In the typical transgenic animals described herein, the transgene causes cells to express a recombinant form of one of a protein, *e.g.* either agonistic or antagonistic forms. However, transgenic animals in which the recombinant gene is silent are also contemplated, as for example, the FLP or CRE recombinase dependent constructs described below. Moreover, "transgenic animal" also includes those recombinant animals in which gene disruption of one or more genes is caused by human intervention, including both recombination and antisense techniques.

The term "treatment", or "treating", as used herein, is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a

symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

The term "wild-type allele" refers to an allele of a gene which, when present in two copies in a subject results in a wild-type phenotype. There can be several different wild-type alleles of a specific gene, since certain nucleotide changes in a gene may not affect the phenotype of a subject having two copies of the gene with the nucleotide changes.

Nucleic Acids of the Present Invention

As described below, one aspect of the invention pertains to isolated nucleic acids comprising an intronic sequence of an SR-BI gene. In a preferred embodiment, the invention provides an intronic sequence of the genomic DNA sequence encoding an SR-BI protein, comprising an intronic sequence shown in Figure 2A-G or set forth in any of SEQ ID NOs.1-121 or complements thereof or homologues thereof. Other preferred nucleic acids of the invention include specific SR-BI alleles, which differ from the allelic variant having the nucleotide sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 3, or at least a portion thereof having a polymorphic region. Nucleic acids of the invention can function as probes or

primers, *e.g.*, in methods for determining the identity of an allelic variant of an SR-BI polymorphic region. The nucleic acids of the invention can also be used to determine whether a subject is at risk of developing a disease associated with a specific allelic variant of an SR-BI polymorphic region, e.g, a disease or disorder associated with an aberrant SR-BI activity. The nucleic acids of the invention can further be used to prepare SR-BI polypeptides encoded by specific alleles, such as mutant alleles. Such polypeptides can be used in gene therapy. Polypeptides encoded by specific SR-BI alleles, such as mutant SR-BI polypeptides, can also be used for preparing reagents, *e.g.*, antibodies, for detecting SR-BI proteins encoded by these alleles. Accordingly, such reagents can be used to detect mutant SR-BI proteins.

Certain nucleic acids of the invention comprise an intronic sequence of an SR-BI gene. The term "SR-BI intronic sequence" refers to a nucleotide sequence of an intron of an SR-BI gene. An intronic sequence can be directly adjacent to an exon or located further away from the exons. Preferred nucleic acids of the invention include an intronic sequence of an SR-BI gene which is adjacent to an exon and comprises at least about 3 consecutive nucleotides, at least about 6 consecutive nucleotides, at least about 9 consecutive nucleotides, at least about 12 consecutive nucleotides, at least about 15 consecutive nucleotides, at least about 18 consecutive nucleotides, or at least about 20 consecutive nucleotides. Isolated nucleic acids which comprise an SR-BI intronic sequence which is immediately adjacent to an exon and comprises at least about 25 consecutive nucleotides, at least about 30 consecutive nucleotides, at least about 35 consecutive nucleotides, at least about 40 consecutive nucleotides, at least about 50 consecutive nucleotides, or at least about 100 consecutive nucleotides are also within the scope of the invention. Preferred isolated nucleic acids of the invention also include those having an SR-BI intronic sequence having a nucleotide sequence of at least about 10 nucleotides, at least about 15 nucleotides, at least about 20 nucleotides, at least about 25 nucleotides, at least about 30 nucleotides, at least about 35 nucleotides, at least about 40 nucleotides, at least about 50 nucleotides or at least about 100 nucleotides. Other preferred nucleic acids of the invention can comprise an SR-BI intronic sequence having less than about 10 nucleotides, provided that the nucleotide sequence is novel. Yet other preferred isolated nucleic acids of the invention include SR-BI intronic nucleic acid sequences of an SR-BI intron, having at least about 150 consecutive nucleotides, at least about 200 consecutive nucleotides, at least about 250 consecutive nucleotides, at least about 300

consecutive nucleotides, at least about 350 consecutive nucleotides, at least about 400 consecutive nucleotides, at least about 500 consecutive nucleotides or at least about 1000 consecutive nucleotides

Preferred nucleic acids of the invention comprise an SR-BI intronic sequence having a nucleotide sequence shown in Figure 2A-G, and/or in any of SEQ ID Nos. 1-121, complement thereof, reverse complement thereof or homologue thereof. In a preferred embodiment, the invention provides an isolated nucleic acid comprising an SR-BI intronic which is at least about 70% 75%, 80%, 85%, 90%, 95%, or preferably at least about 98%, and most preferably at least about 99% identical to an intronic nucleotide sequence shown in Figure 2A-G or set forth in any of SEQ ID NOS.1-121 or a complement thereof. In fact, as described herein, several alleles of human SR-BI genes have been identified. The invention is intended to encompass all of these alleles and SR-BI alleles not yet identified, which can be identified, e.g, according to the methods described herein.

The invention also provides isolated nucleic acids comprising at least one polymorphic region of an SR-BI gene having a nucleotide sequence which differs from the nucleotide sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 3. Preferred nucleic acids have a polymorphic region located in an exon of an SR-BI gene, such as exons 1, 3 or 8. Accordingly, preferred nucleic acids of the invention comprise an adenine at position 146 of exon 1 (as set forth in SEQ ID NO: 95), an adenine at position 119 of exon 3 (as set forth in SEQ ID NO: 96, and/or a thymidine at position 41 of exon 8 (as set forth in SEQ ID NO: 65). Preferred nucleic acids can also have a polymorphic region in an intron, *e.g.*, intron 5 or 10. For example, the invention provides nucleic acids having a polymorphic nucleotide at position 54 of intron 5 and/or at position -41 of intron 10. In a preferred embodiment, the nucleic acid has a thymidine at position 54 of intron 5 (as set forth in SEQ ID NO: 66) and/or a guanine at position -41 of intron 10 (as set forth in SEQ ID NO: 97). The nucleic acids can be genomic DNA, cDNA, or RNA (in which case, the nucleic acid has a uridine at position 54 of intron 5).

Also within the scope of the invention are isolated nucleic acids which encode an SR-BI protein, such as an SR-BI protein having an amino acid sequence which differs from the amino acid sequence set forth in SEQ ID NOs 2 and 4. Preferred nucleic acids encode an SR-

BI polypeptide comprising an amino acid sequence from SEQ ID NO: 2 or 4 in which residue 2 is a serine and/or in which residue 135 is an isoleucine.

Preferred nucleic acids of the invention are from vertebrate genes encoding SR-BI proteins. Particularly preferred vertebrate nucleic acids are mammalian nucleic acids. A particularly preferred nucleic acid of the invention is a human nucleic acid, such as a nucleic acid comprising an SR-BI intronic sequence shown in Figure 2A-G or set forth in any of SEQ ID NOS. 1-121 or complement thereof or an allele comprising a nucleotide sequence set forth in SEQ ID NO: 65 or SEQ ID NO: 97.

Another aspect of the invention provides a nucleic acid which hybridizes under appropriate stringency to an SR-BI intronic sequence having a nucleotide sequence shown in introns shown in Figure 2A-G or in intronic sequences set forth in any of SEQ ID Nos. 1-121 or complement thereof. Appropriate stringency conditions which promote DNA hybridization, for example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 x SSC at 50°C to a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C. Both temperature and salt may be varied, or temperature or salt concentration may be held constant while the other variable is changed. In a preferred embodiment, a nucleic acid of the present invention will bind to at least about 20, preferably at least about 25, more preferably at least about 30 and most preferably at least about 50 consecutive nucleotides of a sequence shown in Figure 2A-G or set forth in any of SEQ ID Nos.1-121 under moderately stringent conditions, for example at about 2.0 x SSC and about 40°C. Even more preferred nucleic acids of the invention are capable of hybridizing under stringent conditions to an intronic sequence of at least about 20, 30, 40, or at least about 50 nucleotides as shown in Figure 2A-G or as set forth in an intronic sequence of any of SEQ ID Nos.1-121.

Hybridization, as described above, can be used to isolate nucleic acids comprising an SR-BI intron or portion thereof from various animal species. A comparison of these nucleic acids should be indicative of intronic sequences which may have a regulatory or other

function, since these regions are expected to be conserved among various species. Hybridization can also be used to isolate SR-BI alleles.

The nucleic acid of the invention can be single stranded DNA (*e.g.*, an oligonucleotide), double stranded DNA (*e.g.*, double stranded oligonucleotide) or RNA. Preferred nucleic acids of the invention can be used as probes or primers. Primers of the invention refer to nucleic acids which hybridize to a nucleic acid sequence which is adjacent to the region of interest or which covers the region of interest and is extended. A primer can be used alone in a detection method, or a primer can be used together with at least one other primer or probe in a detection method. Primers can also be used to amplify at least a portion of a nucleic acid. Probes of the invention refer to nucleic acids which hybridize to the region of interest and which are not further extended. For example, a probe is a nucleic acid which hybridizes to a polymorphic region of an SR-BI gene, and which by hybridization or absence of hybridization to the DNA of a subject will be indicative of the identity of the allelic variant of the polymorphic region of the SR-BI gene.

Numerous procedures for determining the nucleotide sequence of a nucleic acid, or for determining the presence of mutations in nucleic acids include a nucleic acid amplification step, which can be carried out by, *e.g.*, polymerase chain reaction (PCR). Accordingly, in one embodiment, the invention provides primers for amplifying portions of an SR-BI gene, such as portions of exons and/or portions of introns. In a preferred embodiment, the exons and/or sequences adjacent to the exons of the human SR-BI gene will be amplified to, *e.g.*, detect which allelic variant of a polymorphic region is present in the SR-BI gene of a subject. Preferred primers comprise a nucleotide sequence complementary to an SR-BI intronic sequence or a specific allelic variant of an SR-BI polymorphic region and of sufficient length to selectively hybridize with an SR-BI gene. In a preferred embodiment, the primer, *e.g.*, a substantially purified oligonucleotide, comprises a region having a nucleotide sequence which hybridizes under stringent conditions to about 6, 8, 10, or 12, preferably 25, 30, 40, 50, or 75 consecutive nucleotides of an SR-BI gene. In an even more preferred embodiment, the primer is capable of hybridizing to an SR-BI intron and has a nucleotide sequence of an intronic sequence shown in Figure 2A-G or set forth in any of SEQ ID Nos. 1-121, complements thereof, allelic variants thereof, or complements of allelic variants thereof. For example, primers comprising a nucleotide sequence of at least about 15 consecutive

nucleotides, at least about 20 nucleotides or having from about 15 to about 25 nucleotides shown in Figure 2A-G or set forth in any of SEQ ID NOS. 1-121 or complement thereof are provided by the invention. Primers having a sequence of more than about 25 nucleotides are also within the scope of the invention. Preferred primers of the invention are primers that can be used in PCR for amplifying each of the exons of an SR-BI gene. Even more preferred primers of the invention have the nucleotide sequence set forth in any of SEQ ID Nos. 41-64 and 89-94 (see Table VII and X in the Examples).

Primers can be complementary to nucleotide sequences located close to each other or further apart, depending on the use of the amplified DNA. For example, primers can be chosen such that they amplify DNA fragments of at least about 10 nucleotides or as much as several kilobases. Preferably, the primers of the invention will hybridize selectively to nucleotide sequences located about 150 to about 350 nucleotides apart.

For amplifying at least a portion of a nucleic acid, a forward primer (i.e., 5' primer) and a reverse primer (i.e., 3' primer) will preferably be used. Forward and reverse primers hybridize to complementary strands of a double stranded nucleic acid, such that upon extension from each primer, a double stranded nucleic acid is amplified. A forward primer can be a primer having a nucleotide sequence or a portion of the nucleotide sequence shown in Figure 2A-G or in SEQ ID Nos. 1-40, 65, 66, and 95-97. A reverse primer can be a primer having a nucleotide sequence or a portion of the nucleotide sequence that is complementary to a nucleotide sequence shown in Figure 2A-G or in SEQ ID Nos. 1-40, 65, 66, and 95-97. Preferred forward primers comprise a nucleotide sequence set forth in SEQ ID Nos. 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, and 85 (shown in Table VII). Preferred reverse primers comprise a nucleotide sequence set forth in SEQ ID Nos. 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, and 86. Preferred pairs of primers for amplifying each of the exons of human SR-BI are set forth in Table VII.

Yet other preferred primers of the invention are nucleic acids which are capable of selectively hybridizing to an allelic variant of a polymorphic region of an SR-BI gene. Thus, such primers can be specific for an SR-BI gene sequence, so long as they have a nucleotide sequence which is capable of hybridizing to an SR-BI gene. Preferred primers are capable of specifically hybridizing to an allelic variant in which nucleotide 146 of exon 1 of human SR-BI is an adenine, *e.g.*, a nucleic acid having SEQ ID NO: 95; an allelic variant in which

nucleotide 119 of exon 3 is an adenine, *e.g.*, a nucleic acid having SEQ ID NO: 96; or an allelic variant in which nucleotide 41 of exon 8 of human SR-BI is a thymidine, *e.g.*, a nucleic acid having SEQ ID NO: 65. Other preferred primers are capable of specifically hybridizing to an allelic variant in which nucleotide 54 of intron 5 is a thymidine, *e.g.*, a nucleic acid having SEQ ID NO: 66 or nucleotide -41 of intron 10 is a guanine, *e.g.*, a nucleic acid having SEQ ID NO: 97. Such primers can be used, *e.g.*, in sequence specific oligonucleotide priming as described further herein.

The SR-BI nucleic acids of the invention can also be used as probes, *e.g.*, in therapeutic and diagnostic assays. For instance, the present invention provides a probe comprising a substantially purified oligonucleotide, which oligonucleotide comprises a region having a nucleotide sequence that hybridizes under stringent conditions to at least approximately 6, 8, 10 or 12, preferably about 25, 30, 40, 50 or 75 consecutive nucleotides of an SR-BI gene. In one embodiment, the probes preferably hybridize to an intron of an SR-BI gene, having an intronic nucleotide sequence shown in Figure 2A-G or set forth in any of SEQ ID Nos. 1-121, allelic variants thereof, complements thereof or complements of allelic variants thereof. In another embodiment, the probes are capable of hybridizing to a nucleotide sequence encompassing an intron/exon border of an SR-BI gene.

Other preferred probes of the invention are capable of hybridizing specifically to a region of an SR-BI gene which is polymorphic. In an even more preferred embodiment of the invention, the probes are capable of hybridizing specifically to one allelic variant of an SR-BI gene having a nucleotide sequence which differs from the nucleotide sequence set forth in SEQ ID NO: 1 or 3. Such probes can then be used to specifically detect which allelic variant of a polymorphic region of an SR-BI gene is present in a subject. The polymorphic region can be located in the promoter, exon, or intron sequences of an SR-BI gene.

For example, preferred probes of the invention are capable of hybridizing specifically to a region overlapping nucleotide 146 of exon 1 of the human SR-BI gene. In one embodiment, the probe overlapping nucleotide 146 of exon 1 is capable of hybridizing specifically to a nucleotide sequence wherein nucleotide 146 is an adenine (as shown in SEQ ID NO: 95). Examples of such probes include a probe having the nucleotide sequence 5' GCGGAGCAGCTCATGTCTGCG 3' (SEQ ID NO: 98); 5' CTTTCGCGGAGCAGCTCATGTCTGCGCGCCT 3' (SEQ ID NO: 99); and probes having

the complement of these nucleotide sequences, i.e., 5' CGCAGACATGAGCTGCTCCGC 3' (SEQ ID NO: 100); 5' AGGCGCGCAGACATGAGCTGCTCCGCCAAAG 3' (SEQ ID NO: 101). The bold nucleotides represents the location of the nucleotide polymorphism. In another embodiment, the probe overlapping nucleotide 146 of exon 1 is capable of specifically hybridizing to a nucleotide sequence wherein nucleotide 146 is a guanine (as shown in Figure 2A-G and set forth in SEQ ID NO: 5). Examples of such probes include a probe having the nucleotide sequence 5' GCGGAGCAGC**G**CATGTCTGCG 3' (SEQ ID NO: 102); CTTTCGCGGAGCAGC**G**CATGTCTGCGCGCCT 3' (SEQ ID NO: 103) and probes having the complement of these nucleotide sequences, i.e., 5' CGCAGACATGCGCTGCTCCGC 3' (SEQ ID NO: 104); 5' AGGCGCGCAGACATGCGCTGCTCCGCCAAAG 3' (SEQ ID NO: 105).

Preferred probes of the invention are capable of hybridizing specifically to a region overlapping nucleotide 119 of exon 3 of the human SR-BI gene. In one embodiment, the probe overlapping nucleotide 119 of exon 3 is capable of hybridizing specifically to a nucleotide sequence wherein nucleotide 119 is an adenine (as shown in SEQ ID NO: 96). Examples of such probes include a probe having the nucleotide sequence 5' TTGGGCATGATGATGTAGACG 3' (SEQ ID NO: 106); 5' GGATGTTGGGCATGATGATGTAGACGCTCTC 3' (SEQ ID NO: 107); and probes having the complement of these nucleotide sequences, i.e., 5' CGACTACATCATCATGCCCAA 3' (SEQ ID NO: 108); 5' GAGAGCGACTACATCATCATGCCCAACATCC 3' (SEQ ID NO: 109). The bold nucleotides represents the location of the nucleotide polymorphism. In another embodiment, the probe overlapping nucleotide 119 of exon 3 is capable of specifically hybridizing to a nucleotide sequence wherein nucleotide 119 is a guanine (as shown in Figure 2A-G and set forth in SEQ ID NO: 7). Examples of such probes include a probe having the nucleotide sequence 5' TTGGGCATGAGGATGTAGACG 3' (SEQ ID NO: 110); GGATGTTGGGCATGAGGATGTAGACGCTCTC 3' (SEQ ID NO: 111) and probes having the complement of these nucleotide sequences, i.e., 5' CGACTACATCCTCATGCCCAA 3' (SEQ ID NO: 112); 5' GAGAGCGACTACATCCATCATGCCCAACATCC 3' (SEQ ID NO: 113).

Other preferred probes of the invention are capable of hybridizing specifically to a region overlapping nucleotide 41 of exon 8 of the human SR-BI gene. In one embodiment,

the probe overlapping nucleotide 41 of exon 8 is capable of hybridizing specifically to a nucleotide sequence wherein nucleotide 41 is a thymidine (as shown in SEQ ID NO: 65). Examples of such probes include a probe having the nucleotide sequence 5' AACCGGGTC**A**GCGTTGAGGA 3' (SEQ ID NO: 67); 5' TGCCAGAACCGGGTC**A**GCGTTGAGGAAGTGA 3' (SEQ ID NO: 68); and probes having the complement of these nucleotide sequences, i.e., 5'TCCTCAACGC**T**GACCCGGTT 3' (SEQ ID NO: 69); 5' TCACTTCCTCAACGC**T**GACCCGGTTCTGGCA 3' (SEQ ID NO: 70). The bold nucleotides represents the location of the nucleotide polymorphism. In another embodiment, the probe overlapping nucleotide 41 of exon 8 is capable of specifically hybridizing to a nucleotide sequence wherein nucleotide 41 is a cytidine (as shown in Figure 2A-G and set forth in SEQ ID NO: 12). Examples of such probes include a probe having the nucleotide sequence 5' AACCGGGTC**G**GCGTTGATGA 3' (SEQ ID NO: 71); TGCCAGAACCGGGTC**G**GCGTTGATGAAGTGA 3' (SEQ ID NO: 72) and probes having the complement of these nucleotide sequences, i.e., 5' TCATCAACGC**C**GACCCGGTT 3' (SEQ ID NO: 73); 5' TCACTTCATCAACGC**C**GACCCGGTTCTGGCA 3' (SEQ ID NO: 74).

Yet other preferred probes of the invention are capable of hybridizing specifically to a region overlapping nucleotide 54 of intron 5 of the human SR-BI gene. In one embodiment, the probe overlapping nucleotide 54 of intron 5 is capable of hybridizing specifically to a nucleotide sequence wherein nucleotide 54 is a cytidine (as shown in Figure 2A-G and set forth in SEQ ID NOS. 9 and 26). Examples of such probes include a probe having the nucleotide sequence 5' AGCCATGGCC**G**GGCCCACCCT 3' (SEQ ID NO: 75); 5' CGAGCAGCCATGGCC**G**GGCCCACCC**T**CCCCT 3' (SEQ ID NO: 76) ; and probes having the complement of these nucleotide sequences, i.e., 5' AGGGTGGGCC**C**GGCCATGGCT 3' (SEQ ID NO: 77); 5' AGGGGAGGGTGGGCC**C**GGCCATGGCTGCTCG 3' (SEQ ID NO: 78). In another embodiment, the probe overlapping nucleotide 54 of intron 5 is capable of specifically hybridizing to a nucleotide sequence wherein nucleotide 54 is a thymidine (as shown in SEQ ID NO: 66). Examples of such probes include a probe having the nucleotide sequence 5' AGCCATGGCC**A**GGCCCACCCT 3' (SEQ ID NO: 79); 5' CGAGCAGCCATGGCC**A**GGCCCACCCTCCCCT 3' (SEQ ID NO: 80) ; and probes having the complement of these nucleotide sequences, i.e., 5' AGGGTGGGCC**T**GGCCATGGCT 3'

(SEQ ID NO: 81); 5' AGGGGAGGGTGGGCC**T**GGCCATGGCTGCTCG 3' (SEQ ID NO: 82).

Still other preferred probes of the invention are capable of hybridizing specifically to a region overlapping nucleotide -41 of intron 10 of the human SR-BI gene. In one embodiment, the probe overlapping nucleotide -41 of intron 10 is capable of hybridizing specifically to a nucleotide sequence wherein nucleotide -41 is a guanine (as shown in SEQ ID NO: 97). Examples of such probes include a probe having the nucleotide sequence 5' TGGGGCCGCA**C**GCTGCGGGCT 3' (SEQ ID NO: 114); 5' TGAGCTGGGGCCGCA**C**GCTGCGGGCTACAGC 3' (SEQ ID NO: 115); and probes having the complement of these nucleotide sequences, i.e., 5' AGCCCGCAGCGTGCGGCCCCA 3' (SEQ ID NO: 116); 5' GCTGTAGCCCGCAGCGTGCGGCCCCAGCTCA 3' (SEQ ID NO: 117). The bold nucleotides represents the location of the nucleotide polymorphism. In another embodiment, the probe overlapping nucleotide -41 of intron 10 is capable of specifically hybridizing to a nucleotide sequence wherein nucleotide -41 is a cytidine (as shown in Figure 2A-G and set forth in SEQ ID NO: 15). Examples of such probes include a probe having the nucleotide sequence 5' TGGGGCCGCAGGCTGCGGGCT 3' (SEQ ID NO: 118); TGAGCTGGGGCCGCAGGCTGCGGGCTACAGC 3' (SEQ ID NO: 119) and probes having the complement of these nucleotide sequences, i.e., 5' AGCCCGCAGCCTGCGGCCCCA 3' (SEQ ID NO: 120); 5' GCTGTAGCCCGCAGCCTGCGGCCCCAGCTCA 3' (SEQ ID NO: 121).

Particularly, preferred probes of the invention have a number of nucleotides sufficient to allow specific hybridization to the target nucleotide sequence. Where the target nucleotide sequence is present in a large fragment of DNA, such as a genomic DNA fragment of several tens or hundreds of kilobases, the size of the probe may have to be longer to provide sufficiently specific hybridization, as compared to a probe which is used to detect a target sequence which is present in a shorter fragment of DNA. For example, in some diagnostic methods, a portion of an SR-BI gene may first be amplified and thus isolated from the rest of the chromosomal DNA and then hybridized to a probe. In such a situation, a shorter probe will likely provide sufficient specificity of hybridization. For example, a probe having a nucleotide sequence of about 10 nucleotides may be sufficient.

In preferred embodiments, the probe or primer further comprises a label attached thereto, which, *e.g.*, is capable of being detected, *e.g.* the label group is selected from amongst radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors.

In a preferred embodiment of the invention, the isolated nucleic acid, which is used, *e.g.*, as a probe or a primer, is modified, such as to become more stable. Exemplary nucleic acid molecules which are modified include phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patents 5,176,996; 5,264,564; and 5,256,775).

The nucleic acids of the invention can also be modified at the base moiety, sugar moiety, or phosphate backbone, for example, to improve stability of the molecule. The nucleic acids, *e.g.*, probes or primers, may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger et al., 1989, Proc. Natl. Acad. Sci. U.S.A. 86:6553-6556; Lemaitre et al., 1987, Proc. Natl. Acad. Sci. 84:648-652; PCT Publication No. WO88/09810, published December 15, 1988), hybridization-triggered cleavage agents. (See, *e.g.*, Krol et al., 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, Pharm. Res. 5:539-549). To this end, the nucleic acid of the invention may be conjugated to another molecule, *e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

The isolated nucleic acid comprising an SR-BI intronic sequence may comprise at least one modified base moiety which is selected from the group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytidine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytidine, 5-methylcytidine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v),.wybutoxosine, pseudouracil, queosine, 2-thiocytidine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-

oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

The isolated nucleic acid may also comprise at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the nucleic acid comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

In yet a further embodiment, the nucleic acid is an α-anomeric oligonucleotide. An α-anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gautier et al., 1987, *Nucl. Acids Res.* 15:6625-6641). The oligonucleotide is a 2'-0-methylribonucleotide (Inoue et al., 1987, *Nucl. Acids Res.* 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue et al., 1987, *FEBS Lett.* 215:327-330).

Any nucleic acid fragment of the invention can be prepared according to methods well known in the art and described, *e.g.*, in Sambrook, J. Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. For example, discrete fragments of the DNA can be prepared and cloned using restriction enzymes. Alternatively, discrete fragments can be prepared using the Polymerase Chain Reaction (PCR) using primers having an appropriate sequence.

Oligonucleotides of the invention may be synthesized by standard methods known in the art, *e.g.* by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate oligonucleotides may be synthesized by the method of Stein et al. (1988, Nucl. Acids Res. 16:3209), methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7448-7451), etc.

The invention also provides vectors and plasmids containing the nucleic acids of the invention. For example, in one embodiment, the invention provides a vector comprising at least a portion of an SR-BI gene comprising a polymorphic region and/or intronic sequence. Thus, the invention provides vectors for expressing at least a portion of the newly identified

allelic variants of the human SR-BI gene, as well as other allelic variants, having a nucleotide sequence which is different from the nucleotide sequence disclosed in Calvo and Vega, *supra*. The allelic variants can be expressed in eukaryotic cells, *e.g.*, cells of a subject, or in prokaryotic cells.

In one embodiment, the vector comprising at least a portion of an SR-BI allele is introduced into a host cell, such that a protein encoded by the allele is synthesized. The SR-BI protein produced can be used, *e.g.*, for the production of antibodies, which can be used, *e.g.*, in methods for detecting mutant forms of SR-BI. Alternatively, the vector can be used for gene therapy, and be, *e.g.*, introduced into a subject to produce SR-BI protein. Host cells comprising a vector having at least a portion of an SR-BI gene are also within the scope of the invention.

Polypeptides of the present invention

The present invention makes available isolated SR-BI polypeptides, such as SR-BI polypeptides which are encoded by specific allelic variants of SR-BI, such as those identified herein. Accordingly, preferred SR-BI polypeptides of the invention have an amino acid sequence which differs from SEQ ID NOs. 2 and 4. In one embodiment, the SR-BI polypeptides are isolated from, or otherwise substantially free of other cellular proteins. The term "substantially free of other cellular proteins" (also referred to herein as "contaminating proteins") or "substantially pure or purified preparations" are defined as encompassing preparations of SR-BI polypeptides having less than about 20% (by dry weight) contaminating protein, and preferably having less than about 5% contaminating protein. Functional forms of the subject polypeptides can be prepared, for the first time, as purified preparations by using a cloned gene as described herein.

Preferred SR-BI proteins of the invention have an amino acid sequence which is at least about 60%, 70%, 80%, 85%, 90%, or 95% identical or homologous to an amino acid sequence of SEQ ID NOs. 2 or 4. Even more preferred SR-BI proteins comprise an amino acid sequence which is at least about 97, 98, or 99% homologous or identical to an amino acid sequence of SEQ ID NO. 2 or 4. Such proteins can be recombinant proteins, and can be, *e.g.*, produced *in vitro* from nucleic acids comprising a specific allele of an SR-BI polymorphic region. For example, recombinant polypeptides preferred by the present invention can be encoded by a nucleic acid, which is at least 85% homologous and more preferably 90%

homologous and most preferably 95% homologous with a nucleotide sequence set forth in SEQ ID NOS. 1 or 3, and comprises an allele of a polymorphic region that differs from that set forth in SEQ ID NOs. 1 and 3. Polypeptides which are encoded by a nucleic acid that is at least about 98-99% homologous with the sequence of SEQ ID NOs: 1 or 3 and comprises an allele of a polymorphic region that differs from that set forth in SEQ ID NOs. 1 and 3 are also within the scope of the invention.

In a preferred embodiment, an SR-BI protein of the present invention is a mammalian SR-BI protein. In an even more preferred embodiment, the SR-BI protein is a human protein, such as an SR-BI polypeptide comprising an amino acid sequence from SEQ ID NO.2 in which amino acid 2 is a serine and/or amino acid 135 is an isoleucine. Other preferred SR-BI polypeptides comprise an amino acid sequence from SEQ ID NO: 4 in which amino acid 2 is a serine.

SR-BI polypeptides preferably are capable of functioning in one of either role of an agonist or antagonist of at least one biological activity of a wild-type ("authentic") SR-BI protein of the appended sequence listing. The term "evolutionarily related to", with respect to amino acid sequences of SR-BI proteins, refers to both polypeptides having amino acid sequences which have arisen naturally, and also to mutational variants of human SR-BI polypeptides which are derived, for example, by combinatorial mutagenesis.

Full length proteins or fragments corresponding to one or more particular motifs and/or domains or to arbitrary sizes, for example, at least 5, 10, 25, 50, 75 and 100, amino acids in length are within the scope of the present invention.

Isolated peptidyl portions of SR-BI proteins can be obtained by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such peptides. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, an SR-BI polypeptide of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments which can function as either agonists or antagonists of a wild-type (*e.g.*, "authentic") SR-BI protein.

In general, polypeptides referred to herein as having an activity (*e.g.*, are "bioactive") of an SR-BI protein are defined as polypeptides which mimic or antagonize all or a portion

of the biological/biochemical activities of an SR-BI protein having SEQ ID NOs 2 or 4, such as the ability to bind lipids. Other biological activities of the subject SR-BI proteins are described herein or will be reasonably apparent to those skilled in the art. According to the present invention, a polypeptide has biological activity if it is a specific agonist or antagonist of a naturally-occurring form of an SR-BI protein.

Assays for determining whether an SR-BI protein or variant thereof, has one or more biological activities are well known in the art.

Other preferred proteins of the invention are those encoded by the nucleic acids set forth in the section pertaining to nucleic acids of the invention. In particular, the invention provides fusion proteins, *e.g.*, SR-BI-immunoglobulin fusion proteins. Such fusion proteins can provide, *e.g.*, enhanced stability and solubility of SR-BI proteins and may thus be useful in therapy. Fusion proteins can also be used to produce an immunogenic fragment of an SR-BI protein. For example, the VP6 capsid protein of rotavirus can be used as an immunologic carrier protein for portions of the SR-BI polypeptide, either in the monomeric form or in the form of a viral particle. The nucleic acid sequences corresponding to the portion of a subject SR-BI protein to which antibodies are to be raised can be incorporated into a fusion gene construct which includes coding sequences for a late vaccinia virus structural protein to produce a set of recombinant viruses expressing fusion proteins comprising SR-BI epitopes as part of the virion. It has been demonstrated with the use of immunogenic fusion proteins utilizing the Hepatitis B surface antigen fusion proteins that recombinant Hepatitis B virions can be utilized in this role as well. Similarly, chimeric constructs coding for fusion proteins containing a portion of an SR-BI protein and the poliovirus capsid protein can be created to enhance immunogenicity of the set of polypeptide antigens (see, for example, EP Publication No: 0259149; and Evans et al. (1989) Nature 339:385; Huang et al. (1988) J. Virol. 62:3855; and Schlienger et al. (1992) J. Virol. 66:2).

The Multiple antigen peptide system for peptide-based immunization can also be utilized to generate an immunogen, wherein a desired portion of an SR-BI polypeptide is obtained directly from organo-chemical synthesis of the peptide onto an oligomeric branching lysine core (see, for example, Posnett et al. (1988) JBC 263:1719 and Nardelli et al. (1992) J. Immunol. 148:914). Antigenic determinants of SR-BI proteins can also be expressed and presented by bacterial cells.

In addition to utilizing fusion proteins to enhance immunogenicity, it is widely appreciated that fusion proteins can also facilitate the expression of proteins, and accordingly, can be used in the expression of the SR-BI polypeptides of the present invention. For example, SR-BI polypeptides can be generated as glutathione-S-transferase (GST-fusion) proteins. Such GST-fusion proteins can enable easy purification of the SR-BI polypeptide, as for example by the use of glutathione-derivatized matrices (see, for example, Current Protocols in Molecular Biology, eds. Ausubel et al. (N.Y.: John Wiley & Sons, 1991)).

The present invention further pertains to methods of producing the subject SR-BI polypeptides. For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding the subject polypeptides can be cultured under appropriate conditions to allow expression of the peptide to occur. Suitable media for cell culture are well known in the art. The recombinant SR-BI polypeptide can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such peptide. In a preferred embodiment, the recombinant SR-BI polypeptide is a fusion protein containing a domain which facilitates its purification, such as GST fusion protein.

Moreover, it will be generally appreciated that, under certain circumstances, it may be advantageous to provide homologues of one of the subject SR-BI polypeptides which function in a limited capacity as one of either an SR-BI agonist (mimetic) or an SR-BI antagonist, in order to promote or inhibit only a subset of the biological activities of the naturally-occurring form of the protein. Thus, specific biological effects can be elicited by treatment with a homologue of limited function, and with fewer side effects relative to treatment with agonists or antagonists which are directed to all of the biological activities of naturally occurring forms of SR-BI proteins.

Homologues of each of the subject SR-BI proteins can be generated by mutagenesis, such as by discrete point mutation(s), or by truncation. For instance, mutation can give rise to homologues which retain substantially the same, or merely a subset, of the biological activity of the SR-BI polypeptide from which it was derived. Alternatively, antagonistic forms of the protein can be generated which are able to inhibit the function of the naturally occurring form of the protein, such as by competitively binding to an SR-BI receptor.

The recombinant SR-BI polypeptides of the present invention also include homologues of SR-BI polypeptides which differ from the SR-BI proteins having SEQ ID NO.2 or 4, such as versions of those protein which are resistant to proteolytic cleavage, as for example, due to mutations which alter ubiquitination or other enzymatic targeting associated with the protein.

SR-BI polypeptides may also be chemically modified to create SR-BI derivatives by forming covalent or aggregate conjugates with other chemical moieties, such as glycosyl groups, lipids, phosphate, acetyl groups and the like. Covalent derivatives of SR-BI proteins can be prepared by linking the chemical moieties to functional groups on amino acid sidechains of the protein or at the N-terminus or at the C-terminus of the polypeptide.

Modification of the structure of the subject SR-BI polypeptides can be for such purposes as enhancing therapeutic or prophylactic efficacy, stability (*e.g.*, ex vivo shelf life and resistance to proteolytic degradation), or post-translational modifications (*e.g.*, to alter phosphorylation pattern of protein). Such modified peptides, when designed to retain at least one activity of the naturally-occurring form of the protein, or to produce specific antagonists thereof, are considered functional equivalents of the SR-BI polypeptides described in more detail herein. Such modified peptides can be produced, for instance, by amino acid substitution, deletion, or addition. The substitutional variant may be a substituted conserved amino acid or a substituted non-conserved amino acid.

For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (i.e. isosteric and/or isoelectric mutations) will not have a major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids can be divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine histidine, (3) aliphatic = glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide =

asparagine, glutamine; and (6) sulfur -containing = cysteine and methionine. (see, for example, Biochemistry, 2nd ed., Ed. by L. Stryer, WH Freeman and Co.: 1981). Whether a change in the amino acid sequence of a peptide results in a functional SR-BI homologue (*e.g.*, functional in the sense that the resulting polypeptide mimics or antagonizes the wild-type form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type protein, or competitively inhibit such a response. Polypeptides in which more than one replacement has taken place can readily be tested in the same manner.

Kits

As set forth herein, the invention provides methods, *e.g.*, diagnostic and therapeutic methods, *e.g.*, for determining the type of allelic variant of a polymorphic region present in an SR-BI gene, such as a human SR-BI gene (*e.g.*,. EX8 or IVS5). In preferred embodiments, the methods use probes or primers comprising nucleotide sequences which are complementary to an SR-BI intronic sequence or to a polymorphic region of an SR-BI gene. Accordingly, the invention provides kits for performing these methods.

In a preferred embodiment, the invention provides a kit for determining whether a subject has or is at risk of developing a disease or condition associated with a specific allelic variant of an SR-BI polymorphic region. In another preferred embodiment, the invention provides a kit for predicting the response (*e.g.*, lowering HDL levels) to HRT in a subject based on the presence or absence of an SR-B1 variant (*e.g.*. EX8 or IVS5). In an even more preferred embodiment, the disease or disorder is characterized by an abnormal SR-BI activity. In an even more preferred embodiment, the invention provides a kit for determining whether a subject has or is at risk of developing a cardiovascular disease, *e.g.*, ischemia, restenosis, congestive heart failure, atherosclerosis, aberrant lipid (*e.g.*, cholesterol), lipoprotein (*e.g.*, HDL, LDL) or triglyceride levels, gallstone formation, diabetes, or an abnormal body mass index, e.g, obesity or cachexia. A preferred kit provides reagents for determining whether a female subject is has or is likely to develop high LDL levels or a high BMI or whether a male subject has or is likely to develop low HDL levels. Another preferred kit provides reagents for determining whether a subject (*e.g.*, a male or a female subject) has or is likely to develop low HDL levels.

Preferred kits comprise at least one probe or primer which is capable of specifically hybridizing to an SR-BI sequence or polymorphic region and instructions for use. The kits preferably comprise at least one of the above described nucleic acids, *e.g.*, including nucleic acids hybridizing to an exon/intron border. Preferred kits for amplifying at least a portion of an SR-BI gene, *e.g.*, an exon, comprise two primers, at least one of which is capable of hybridizing to an SR-BI intronic sequence or an allelic variant sequence. Even more preferred kits comprise a pair of primers selected from the group consisting of SEQ ID NO: 41 and SEQ ID NO: 42, SEQ ID NO: 43 and SEQ ID NO: 44, SEQ ID NO: 45 and SEQ ID NO: 46, SEQ ID NO: 47 and SEQ ID NO: 48, SEQ ID NO: 49 and SEQ ID NO: 50, SEQ ID NO: 51 and SEQ ID NO: 52, SEQ ID NO: 53 and SEQ ID NO: 54, SEQ ID NO: 55 and SEQ ID NO: 56, SEQ ID NO: 57 and SEQ ID NO: 58, SEQ ID NO: 59 and SEQ ID NO: 60, SEQ ID NO: 61 and SEQ ID NO: 62, SEQ ID NO:63 and SEQ ID NO: 64, SEQ ID NO: 85 and SEQ ID NO: 86, SEQ ID NO:89 and SEQ ID NO: 90, SEQ ID NO:91 and SEQ ID NO: 92, and SEQ ID NO:93 and SEQ ID NO: 94.

The kits of the invention can also comprise one or more control nucleic acids or reference nucleic acids, such as nucleic acids comprising an SR-BI intronic sequence. For example, a kit can comprise primers for amplifying a polymorphic region of an SR-BI gene and a control DNA corresponding to such an amplified DNA and having the nucleotide sequence of a specific allelic variant. Thus, direct comparison can be performed between the DNA amplified from a subject and the DNA having the nucleotide sequence of a specific allelic variant. In one embodiment, the control nucleic acid comprises at least a portion of an SR-BI gene of an individual, who does not have a cardiovascular disease, aberrant lipid levels, gallstones, or a disease or disorder associated with an aberrant SR-BI activity.

Yet other kits of the invention comprise at least one reagent necessary to perform the assay. For example, the kit can comprise an enzyme. Alternatively the kit can comprise a buffer or any other necessary reagent.

Predictive Medicine

The invention further features predictive medicines, which are based, at least in part, on determination of the identity of SR-BI polymorphic regions which are associated with specific diseases or disorders.

For example, information obtained using the diagnostic assays described herein (alone or in conjunction with information on another genetic defect, which contributes to the same disease) is useful for diagnosing or confirming that a symptomatic subject has an allele of a polymorphic region which is associated with a particular disease or disorder. Alternatively, the information (alone or in conjunction with information on another genetic defect, which contributes to the same disease) can be used prognostically for predicting whether a non-symptomatic subject is likely to develop a disease or condition, which is associated with one or more specific alleles of SR-BI polymorphic regions in a subject. Based on the prognostic information, a doctor can recommend a regimen (*e.g.* diet or exercise) or therapeutic protocol, useful for preventing or prolonging onset of the particular disease or condition in the individual.

In addition, knowledge of the identity of a particular SR-BI allele in an individual (the SR-BI genetic profile), alone or in conjunction with information on other genetic defects contributing to the same disease (the genetic profile of the particular disease) allows customization of therapy for a particular disease to the individual's genetic profile, the goal of "pharmacogenomics". For example, an individual's SR-BI genetic profile or the genetic profile of a disease or condition associated with a specific allele of an SR-BI polymorphic region, can enable a doctor: 1) to more effectively prescribe a drug that will address the molecular basis of the disease or condition; and 2) to better determine the appropriate dosage of a particular drug. For example, the expression level of SR-BI proteins, alone or in conjunction with the expression level of other genes, known to contribute to the same disease, can be measured in many patients at various stages of the disease to generate a transcriptional or expression profile of the disease. Expression patterns of individual patients can then be compared to the expression profile of the disease to determine the appropriate drug and dose to administer to the patient.

The ability to target populations expected to show the highest clinical benefit, based on the SR-BI or disease genetic profile, can enable: 1) the repositioning of marketed drugs with disappointing market results; 2) the rescue of drug candidates whose clinical

development has been discontinued as a result of safety or efficacy limitations, which are patient subgroup-specific; and 3) an accelerated and less costly development for drug candidates and more optimal drug labeling (*e.g.* since the use of SR-BI as a marker is useful for optimizing effective dose).

These and other methods are described in further detail in the following sections.

Prognostic and Diagnostic Assays

The present methods provide means for determining if a subject has (diagnostic) or is at risk of developing (prognostic) a disease, condition or disorder that is associated a specific SR-BI allele, *e.g.*, a body mass disorder or an abnormal lipid level (HDL and LDL) and disorders resulting therefrom.

The present invention provides methods for determining the molecular structure of an SR-BI gene, such as a human SR-BI gene, or a portion thereof. In one embodiment, determining the molecular structure of at least a portion of an SR-BI gene comprises determining the identity of the allelic variant of at least one polymorphic region of an SR-BI gene. A polymorphic region of an SR-BI gene can be located in an exon, an intron, at an intron/exon border, or in the promoter of the SR-BI gene.

The invention provides methods for determining whether a subject has, or is at risk of developing, a disease or condition associated with a specific allelic variant of a polymorphic region of an SR-BI gene. Such diseases can be associated with an aberrant SR-BI activity, *e.g.*, abnormal binding to a lipid, or an aberrant SR-BI protein level. An aberrant SR-BI protein level can result from an aberrant transcription or post transcriptional regulation. Thus, allelic differences in specific regions of an SR-BI gene can result in differences of SR-BI protein due to differences in regulation of expression. In particular, some of the identified polymorphisms in the human SR-BI gene may be associated with differences in the level of transcription, RNA maturation, splicing, or translation of the SR-BI gene or transcription product. This invention further provides methods for predicting the effect of HRT on lipid level (*e.g.* HDL or LDL), in females, based on the identification of specific allelic variants of an SR-B1 gene (*e.g.*, EX8 or IVS5).

Analysis of one or more SR-BI polymorphic region in a subject can be useful for predicting whether a subject has or is likely to develop a body mass disorder, an abnormal

lipoprotein or lipid level and disorders resulting therefrom, such as cardiovascular disorders, diabetes and gallstone formation, or for predicting the effect of HRT on lipid level in a subject.

In addition, since SR-BI is a receptor that is capable of binding to various lipid related molecules, it is likely that specific alleles of the SR-BI gene are associated with other diseases or conditions involving an inappropriate lipid transfer or metabolism, *e.g.*, atherosclerosis or a biliary disorder, such as gallstone formation. Accordingly, the invention provides diagnostic and prognostic assays for determining whether a subject is at risk of developing a disease characterized by an abnormal lipid level, e.g, atherosclerosis or gall stone formation.

In preferred embodiments, the methods of the invention can be characterized as comprising detecting, in a sample of cells from the subject, the presence or absence of a specific allelic variant of one or more polymorphic regions of an SR-BI gene. The allelic differences can be: (i) a difference in the identity of at least one nucleotide or (ii) a difference in the number of nucleotides, which difference can be a single nucleotide or several nucleotides. The invention also provides methods for detecting differences in SR-BI genes such as chromosomal rearrangements, *e.g.*, chromosomal dislocation. The invention can also be used in prenatal diagnostics.

A preferred detection method is allele specific hybridization using probes overlapping the polymorphic site and having about 5, 10, 20, 25, or 30 nucleotides around the polymorphic region. Examples of probes for detecting specific allelic variants of the polymorphic region located in exon 1 are probes comprising a nucleotide sequence set forth in any of SEQ ID NO: 98-105; probes for detecting specific allelic variants of the polymorphic region located in exon 3 are probes comprising a nucleotide sequence set forth in any of SEQ ID NO: 106-113; and probes for detecting specific allelic variants of the polymorphic region located in exon 8 are probes comprising a nucleotide sequence set forth in any of SEQ ID NO: 67-74. Examples of probes for detecting specific allelic variants of the polymorphic region located in intron 5 are probes comprising a nucleotide sequence set forth in any of SEQ ID NO: 75-82; and probes for detecting specific allelic variants of the polymorphic region located in intron 10 are probes comprising a nucleotide sequence set forth in any of SEQ ID NO: 114-121. In a preferred embodiment of the invention, several probes capable of hybridizing specifically to allelic variants are attached to a solid phase support,

e.g., a "chip". Oligonucleotides can be bound to a solid support by a variety of processes, including lithography. For example a chip can hold up to 250,000 oligonucleotides (GeneChip, Affymetrix). Mutation detection analysis using these chips comprising oligonucleotides, also termed "DNA probe arrays" is described *e.g.*, in Cronin et al. (1996) Human Mutation 7:244. In one embodiment, a chip comprises all the allelic variants of at least one polymorphic region of a gene. The solid phase support is then contacted with a test nucleic acid and hybridization to the specific probes is detected. Accordingly, the identity of numerous allelic variants of one or more genes can be identified in a simple hybridization experiment. For example, the identity of the allelic variant of the nucleotide polymorphism in exons 1, 3, 8 or in introns 5 and 10 can be determined in a single hybridization experiment.

In other detection methods, it is necessary to first amplify at least a portion of an SR-BI gene prior to identifying the allelic variant. Amplification can be performed, *e.g.*, by PCR and/or LCR, according to methods known in the art. In one embodiment, genomic DNA of a cell is exposed to two PCR primers and amplification for a number of cycles sufficient to produce the required amount of amplified DNA. In preferred embodiments, the primers are located between 150 and 350 base pairs apart. Preferred primers, such as primers for amplifying each of the exons of the human SR-BI gene, are listed in Table IX in the Examples. Details regarding the PCR reaction are indicated in Table VIII, also in the Examples.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. et al., 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. et al., 1989, Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. et al., 1988, Bio/Technology 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In one embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence at least a portion of an SR-BI gene and detect allelic variants, *e.g.*, mutations, by comparing the sequence of the sample sequence with the corresponding wild-type (control) sequence. Exemplary sequencing reactions include those based on techniques

developed by Maxam and Gilbert (*Proc. Natl Acad Sci USA* (1977) 74:560) or Sanger (Sanger et al (1977) *Proc. Nat. Acad. Sci* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures may be utilized when performing the subject assays (*Biotechniques* (1995) 19:448), including sequencing by mass spectrometry (see, for example, U.S. Patent No. 5,547,835 and international patent application Publication Number WO 94/16101, entitled *DNA Sequencing by Mass Spectrometry* by H. Köster; U.S. Patent No. 5,547,835 and international patent application Publication Number WO 94/21822 entitled "DNA Sequencing by Mass Spectrometry Via Exonuclease Degradation" by H. Köster), and U.S Patent No.5,605,798 and International Patent Application No. PCT/US96/03651 entitled *DNA Diagnostics Based on Mass Spectrometry* by H. Köster;. Cohen et al. (1996) *Adv Chromatogr* 36:127-162; and Griffin et al. (1993) *Appl Biochem Biotechnol* 38:147-159). It will be evident to one skilled in the art that, for certain embodiments, the occurrence of only one, two or three of the nucleic acid bases need be determined in the sequencing reaction. For instance, A-track or the like, *e.g.*, where only one nucleotide is detected, can be carried out.

Yet other sequencing methods are disclosed, *e.g.*, in U.S. Patent No. 5,580,732 entitled "Method of DNA sequencing employing a mixed DNA-polymer chain probe" and U.S. Patent No. 5,571,676 entitled "Method for mismatch-directed *in vitro* DNA sequencing".

In some cases, the presence of a specific allele of an SR-BI gene in DNA from a subject can be shown by restriction enzyme analysis. For example, a specific nucleotide polymorphism can result in a nucleotide sequence comprising a restriction site which is absent from the nucleotide sequence of another allelic variant. In particular, the presence of a cytidine at position 54 of intron 5 creates an ApaI site, whereas the presence of a thymidine, at this position destroys the ApaI site. Similarly, the polylmorphism of exon 1 and exon 8 can be determined by analyzing the products or restriction digests (see Table IX).

In a further embodiment, protection from cleavage agents (such as a nuclease, hydroxylamine or osmium tetroxide and with piperidine) can be used to detect mismatched bases in RNA/RNA DNA/DNA, or RNA/DNA heteroduplexes (Myers, et al. (1985) *Science* 230:1242). In general, the technique of "mismatch cleavage" starts by providing heteroduplexes formed by hybridizing a control nucleic acid, which is optionally labeled, *e.g.*, RNA or DNA, comprising a nucleotide sequence of an SR-BI allelic variant with a sample nucleic acid, e.g, RNA or DNA, obtained from a tissue sample. The double-stranded duplexes

are treated with an agent which cleaves single-stranded regions of the duplex such as duplexes formed based on basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digest the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine whether the control and sample nucleic acids have an identical nucleotide sequence or in which nucleotides they are different. See, for example, Cotton et al (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba et al (1992) Methods Enzymod. 217:286-295. In a preferred embodiment, the control or sample nucleic acid is labeled for detection.

In other embodiments, alterations in electrophoretic mobility is used to identify the type of SR-BI allelic variant. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc Natl. Acad. Sci USA* 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; and Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control nucleic acids are denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In another preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) Trends Genet 7:5).

In yet another embodiment, the identity of an allelic variant of a polymorphic region is obtained by analyzing the movement of a nucleic acid comprising the polymorphic region in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al (1985) Nature 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich

DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing agent gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) Biophys Chem 265:1275).

Examples of techniques for detecting differences of at least one nucleotide between 2 nucleic acids include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide probes may be prepared in which the known polymorphic nucleotide is placed centrally (allele-specific probes) and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) *Nature* 324:163); Saiki et al (1989) *Proc. Natl Acad. Sci USA* 86:6230; and Wallace et al. (1979) Nucl. Acids Res. 6:3543). Such allele specific oligonucleotide hybridization techniques may be used for the simultaneous detection of several nucleotide changes in different polylmorphic regions of SR-BI. For example, oligonucleotides having nucleotide sequences of specific allelic variants are attached to a hybridizing membrane and this membrane is then hybridized with labeled sample nucleic acid. Analysis of the hybridization signal will then reveal the identity of the nucleotides of the sample nucleic acid.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the allelic variant of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al. (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238; Newton et al. (1989) Nucl. Acids Res. 17:2503). This technique is also termed "PROBE" for Probe Oligo Base Extension. In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al (1992) *Mol. Cell Probes* 6:1).

In another embodiment, identification of the allelic variant is carried out using an oligonucleotide ligation assay (OLA), as described, *e.g.*, in U.S. Pat. No. 4,998,617 and in Landegren, U. et al., Science 241:1077-1080 (1988). The OLA protocol uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target. One of the oligonucleotides is linked to a separation marker, *e.g.*,

biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate. Ligation then permits the labeled oligonucleotide to be recovered using avidin, or another biotin ligand. Nickerson, D. A. et al. have described a nucleic acid detection assay that combines attributes of PCR and OLA (Nickerson, D. A. et al., Proc. Natl. Acad. Sci. (U.S.A.) 87:8923-8927 (1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Several techniques based on this OLA method have been developed and can be used to detect specific allelic variants of a polymorphic region of an SR-BI gene. For example, U.S. Patent No. 5593826 discloses an OLA using an oligonucleotide having 3'-amino group and a 5'-phosphorylated oligonucleotide to form a conjugate having a phosphoramidate linkage. In another variation of OLA described in Tobe et al. ((1996)Nucleic Acids Res 24: 3728), OLA combined with PCR permits typing of two alleles in a single microtiter well. By marking each of the allele-specific primers with a unique hapten, i.e. digoxigenin and fluorescein, each OLA reaction can be detected by using hapten specific antibodies that are labeled with different enzyme reporters, alkaline phosphatase or horseradish peroxidase. This system permits the detection of the two alleles using a high throughput format that leads to the production of two different colors.

The invention further provides methods for detecting single nucleotide polymorphisms in an SR-BI gene. Because single nucleotide polymorphisms constitute sites of variation flanked by regions of invariant sequence, their analysis requires no more than the determination of the identity of the single nucleotide present at the site of variation and it is unnecessary to determine a complete gene sequence for each patient. Several methods have been developed to facilitate the analysis of such single nucleotide polymorphisms.

In one embodiment, the single base polymorphism can be detected by using a specialized exonuclease-resistant nucleotide, as disclosed, *e.g.*, in Mundy, C. R. (U.S. Pat. No.4,656,127). According to the method, a primer complementary to the allelic sequence immediately 3' to the polymorphic site is permitted to hybridize to a target molecule obtained from a particular animal or human. If the polymorphic site on the target molecule contains a nucleotide that is complementary to the particular exonuclease-resistant nucleotide derivative present, then that derivative will be incorporated onto the end of the hybridized

primer. Such incorporation renders the primer resistant to exonuclease, and thereby permits its detection. Since the identity of the exonuclease-resistant derivative of the sample is known, a finding that the primer has become resistant to exonucleases reveals that the nucleotide present in the polymorphic site of the target molecule was complementary to that of the nucleotide derivative used in the reaction. This method has the advantage that it does not require the determination of large amounts of extraneous sequence data.

In another embodiment of the invention, a solution-based method is used for determining the identity of the nucleotide of a polymorphic site. Cohen, D. et al. (French Patent 2,650,840; PCT Appln. No. WO91/02087). As in the Mundy method of U.S. Pat. No. 4,656,127, a primer is employed that is complementary to allelic sequences immediately 3' to a polymorphic site. The method determines the identity of the nucleotide of that site using labeled dideoxynucleotide derivatives, which, if complementary to the nucleotide of the polymorphic site will become incorporated onto the terminus of the primer.

An alternative method, known as Genetic Bit Analysis or GBA is described by Goelet, P. et al. (PCT Appln. No. 92/15712). The method of Goelet, P. et al. uses mixtures of labeled terminators and a primer that is complementary to the sequence 3' to a polymorphic site. The labeled terminator that is incorporated is thus determined by, and complementary to, the nucleotide present in the polymorphic site of the target molecule being evaluated. In contrast to the method of Cohen et al. (French Patent 2,650,840; PCT Appln. No. WO91/02087) the method of Goelet, P. et al. is preferably a heterogeneous phase assay, in which the primer or the target molecule is immobilized to a solid phase.

Recently, several primer-guided nucleotide incorporation procedures for assaying polymorphic sites in DNA have been described (Komher, J. S. et al., Nucl. Acids. Res. 17:7779-7784 (1989); Sokolov, B. P., Nucl. Acids Res. 18:3671 (1990); Syvanen, A. -C., et al., Genomics 8:684-692 (1990); Kuppuswamy, M. N. et al., Proc. Natl. Acad. Sci. (U.S.A.) 88:1143-1147 (1991); Prezant, T. R. et al., Hum. Mutat. 1:159-164 (1992); Ugozzoli, L. et al., GATA 9:107-112 (1992); Nyren, P. et al., Anal. Biochem. 208:171-175 (1993)). These methods differ from GBA in that they all rely on the incorporation of labeled deoxynucleotides to discriminate between bases at a polymorphic site. In such a format, since the signal is proportional to the number of deoxynucleotides incorporated, polymorphisms

that occur in runs of the same nucleotide can result in signals that are proportional to the length of the run (Syvanen, A. -C., et al., Amer.J. Hum. Genet. 52:46-59 (1993)).

For determining the identity of the allelic variant of a polymorphic region located in the coding region of an SR-BI gene, yet other methods than those described above can be used. For example, identification of an allelic variant which encodes a mutated SR-BI protein can be performed by using an antibody specifically recognizing the mutant protein in, *e.g.*, immunohistochemistry or immunoprecipitation. Antibodies to wild-type SR-BI protein are described, e.g, in Acton et al. (1999) Science 271:518 (anti-mouse SR-BI antibody cross-reactive with human SR-BI). Other antibodies to wild-type SR-BI or mutated forms of SR-BI proteins can be prepared according to methods known in the art. Preferred antibodies specifically bind to a human SR-BI protein having a serine at residue 2 and/or having an isoleucine at amino acid residue 135. Alternatively, one can also measure an activity of an SR-BI protein, such as binding to a lipid or lipoprotein. Binding assays are known in the art and involve, *e.g.*, obtaining cells from a subject, and performing binding experiments with a labeled lipid, to determine whether binding to the mutated form of the receptor differs from binding to the wild-type of the receptor.

Antibodies directed against wild type or mutant SR-BI polypeptides or allelic variant thereof, which are discussed above, may also be used in disease diagnostics and prognostics. Such diagnostic methods, may be used to detect abnormalities in the level of SR-BI polypeptide expression, or abnormalities in the structure and/or tissue, cellular, or subcellular location of an SR-BI polypeptide. Structural differences may include, for example, differences in the size, electronegativity, or antigenicity of the mutant SR-BI polypeptide relative to the normal SR-BI polypeptide. Protein from the tissue or cell type to be analyzed may easily be detected or isolated using techniques which are well known to one of skill in the art, including but not limited to western blot analysis. For a detailed explanation of methods for carrying out Western blot analysis, see Sambrook et al, 1989, supra, at Chapter 18. The protein detection and isolation methods employed herein may also be such as those described in Harlow and Lane, for example, (Harlow, E. and Lane, D., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), which is incorporated herein by reference in its entirety.

This can be accomplished, for example, by immunofluorescence techniques employing a fluorescently labeled antibody (see below) coupled with light microscopic, flow

cytometric, or fluorimetric detection. The antibodies (or fragments thereof) useful in the present invention may, additionally, be employed histologically, as in immunofluorescence or immunoelectron microscopy, for *in situ* detection of SR-BI polypeptides. *In situ* detection may be accomplished by removing a histological specimen from a patient, and applying thereto a labeled antibody of the present invention. The antibody (or fragment) is preferably applied by overlaying the labeled antibody (or fragment) onto a biological sample. Through the use of such a procedure, it is possible to determine not only the presence of the SR-BI polypeptide, but also its distribution in the examined tissue. Using the present invention, one of ordinary skill will readily perceive that any of a wide variety of histological methods (such as staining procedures) can be modified in order to achieve such *in situ* detection.

Often a solid phase support or carrier is used as a support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

One means for labeling an anti-SR-BI polypeptide specific antibody is via linkage to an enzyme and use in an enzyme immunoassay (EIA) (Voller, "The Enzyme Linked Immunosorbent Assay (ELISA)", *Diagnostic Horizons* 2:1-7, 1978, Microbiological Associates Quarterly Publication, Walkersville, MD; Voller, et al., J. Clin. Pathol. 31:507-520 (1978); Butler, Meth. Enzymol. 73:482-523 (1981); Maggio, (ed.) *Enzyme Immunoassay*, CRC Press, Boca Raton, FL, 1980; Ishikawa, et al., (eds.) *Enzyme Immunoassay*, Kgaku Shoin, Tokyo, 1981). The enzyme which is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, fluorimetric or by visual means. Enzymes which can be used to detectably label the antibody include, but

are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods which employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

Detection may also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect fingerprint gene wild type or mutant peptides through the use of a radioimmunoassay (RIA) (see, for example, Weintraub, B., *Principles of Radioimmunoassays*, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March, 1986, which is incorporated by reference herein). The radioactive isotope can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography.

It is also possible to label the antibody with a fluorescent compound. When the fluorescently labeled antibody is exposed to light of the proper wave length, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labeling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine.

The antibody can also be detectably labeled using fluorescence emitting metals such as ^{152}Eu, or others of the lanthanide series. These metals can be attached to the antibody using such metal chelating groups as diethylenetriaminepentacetic acid (DTPA) or ethylenediaminetetraacetic acid (EDTA).

The antibody also can be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of particularly useful chemiluminescent labeling compounds are luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound may be used to label the antibody of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems in, which a catalytic protein increases the efficiency of the chemiluminescent

reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

Moreover, it will be understood that any of the above methods for detecting alterations in a gene or gene product or polymorphic variants can be used to monitor the course of treatment or therapy.

If a polymorphic region is located in an exon, either in a coding or non-coding portion of the gene, the identity of the allelic variant can be determined by determining the molecular structure of the mRNA, pre-mRNA, or cDNA. The molecular structure can be determined using any of the above described methods for determining the molecular structure of the genomic DNA, *e.g.*, sequencing and SSCP.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits, such as those described above, comprising at least one probe or primer nucleic acid described herein, which may be conveniently used, *e.g.*, to determine whether a subject has or is at risk of developing a disease associated with a specific SR-BI allelic variant.

Sample nucleic acid for using in the above-described diagnostic and prognostic methods can be obtained from any cell type or tissue of a subject. For example, a subject's bodily fluid (*e.g.,* blood) can be obtained by known techniques (*e.g.*, venipuncture). Alternatively, nucleic acid tests can be performed on dry samples (*e.g.*, hair or skin). Fetal nucleic acid samples can be obtained from maternal blood as described in International Patent Application No. WO91/07660 to Bianchi. Alternatively, amniocytes or chorionic villi may be obtained for performing prenatal testing.

Diagnostic procedures may also be performed *in situ* directly upon tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections, such that no nucleic acid purification is necessary. Nucleic acid reagents may be used as probes and/or primers for such *in situ* procedures (see, for example, Nuovo, G.J., 1992, PCR *in situ* hybridization: protocols and applications, Raven Press, NY).

In addition to methods which focus primarily on the detection of one nucleic acid sequence, profiles may also be assessed in such detection schemes. Fingerprint profiles may be generated, for example, by utilizing a differential display procedure, Northern analysis and/or RT-PCR.

Pharmacogenomics

Knowledge of the identity of the allele of one or more SR-BI gene polymorphic regions in an individual (the SR-BI genetic profile), alone or in conjunction with information on other genetic defects contributing to the same disease (the genetic profile of the particular disease) allows a customization of the therapy for a particular disease to the individual's genetic profile, the goal of "pharmacogenomics". For example, subjects having a specific allele of an SR-BI gene may or may not exhibit symptoms of a particular disease or be predisposed to developing symptoms of a particular disease. Further, if those subjects are symptomatic, they may or may not respond to a certain drug, *e.g.*, a specific SR-BI therapeutic, but may respond to another. Thus, generation of an SR-BI genetic profile, (*e.g.*, categorization of alterations in SR-BI genes which are associated with the development of a particular disease), from a population of subjects, who are symptomatic for a disease or condition that is caused by or contributed to by a defective and/or deficient SR-BI gene and/or protein (an SR-BI genetic population profile) and comparison of an individual's SR-BI profile to the population profile, permits the selection or design of drugs that are expected to be safe and efficacious for a particular patient or patient population (i.e., a group of patients having the same genetic alteration).

For example, an SR-BI population profile can be performed by determining the SR-BI profile, *e.g.*, the identity of SR-BI alleles, in a patient population having a disease, which is associated with one or more specific alleles of SR-BI polymorphic regions. Optionally, the SR-BI population profile can further include information relating to the response of the population to an SR-BI therapeutic, using any of a variety of methods, including, monitoring: 1) the severity of symptoms associated with the SR-BI related disease, 2) SR-BI gene expression level, 3) SR-BI mRNA level, and/or 4) SR-BI protein level. and (iii) dividing or categorizing the population based on particular SR-BI alleles. The SR-BI genetic population profile can also, optionally, indicate those particular SR-BI alleles which are present in patients that are either responsive or non-responsive to a particular therapeutic. This information or population profile, is then useful for predicting which individuals should respond to particular drugs, based on their individual SR-BI profile.

In a preferred embodiment, the SR-BI profile is a transcriptional or expression level profile and step (i) is comprised of determining the expression level of SR-BI proteins, alone

or in conjunction with the expression level of other genes known to contribute to the same disease at various stages of the disease.

Pharmacogenomic studies can also be performed using transgenic animals. For example, one can produce transgenic mice, *e.g.*, as described herein, which contain a specific allelic variant of an SR-BI gene. These mice can be created, *e.g.,* by replacing their wild-type SR-BI gene with an allele of the human SR-BI gene. The response of these mice to specific SR-BI therapeutics can then be determined.

Monitoring Effects of SR-BI Therapeutics During Clinical Trials

The ability to target populations expected to show the highest clinical benefit, based on the SR-BI or disease genetic profile, can enable: 1) the repositioning of marketed drugs with disappointing market results; 2) the rescue of drug candidates whose clinical development has been discontinued as a result of safety or efficacy limitations, which are patient subgroup-specific; and 3) an accelerated and less costly development for drug candidates and more optimal drug labeling (*e.g.* since the use of SR-BI as a marker is useful for optimizing effective dose).

In situations in which the disease associated with a specific SR-BI allele is characterized by an abnormal SR-BI expression, the treatment of an individual with an SR-BI therapeutic can be monitored by determining SR-BI characteristics, such as SR-BI protein level or activity, SR-BI mRNA level, and/or SR-BI transcriptional level. This measurement will indicate whether the treatment is effective or whether it should be adjusted or optimized. Thus, SR-BI can be used as a marker for the efficacy of a drug during clinical trials.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a preadministration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of an SR-BI protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the SR-BI protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the SR-BI protein, mRNA, or genomic DNA in the preadministration sample with the SR-BI

protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of SR-BI to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of SR-BI to lower levels than detected, i.e., to decrease the effectiveness of the agent.

Cells of a subject may also be obtained before and after administration of an SR-BI therapeutic to detect the level of expression of genes other than SR-BI, to verify that the SR-BI therapeutic does not increase or decrease the expression of genes which could be deleterious. This can be done, *e.g.*, by using the method of transcriptional profiling. Thus, mRNA from cells exposed *in vivo* to an SR-BI therapeutic and mRNA from the same type of cells that were not exposed to the SR-BI therapeutic could be reverse transcribed and hybridized to a chip containing DNA from numerous genes, to thereby compare the expression of genes in cells treated and not treated with an SR-BI therapeutic. If, for example an SR-BI therapeutic turns on the expression of a proto-oncogene in an individual, use of this particular SR-BI therapeutic may be undesirable.

Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject having or likely to develop a disorder associated with specific SR-BI alleles and/or aberrant SR-BI expression or activity, *e.g.*, disorders or diseases associated with an abnormal BMI or lipid levels such as cardiovascular disease or diabetes.

Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with a specific SR-BI allele and/or an aberrant SR-BI expression or activity, such as a body mass disorder or abnormal lipid level and medical conditions resulting therefrom, by administering to the subject an agent which counteracts the unfavorable biological effect of the specific SR-BI allele. Subjects at risk for such a disease can be identified by a diagnostic or prognostic assay, *e.g.*, as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms associated with specific SR-BI alleles, such that a disease or disorder is prevented or, alternatively, delayed in its

progression. Depending on the identity of the SR-BI allele in a subject, a compound that counteracts the effect of this allele is administered. The compound can be a compound modulating the plasma level of lipids. The treatment can also be a specific diet. In particular, the treatment can be undertaken prophylactically, before any other symptoms are present. Such a prophylactic treatment could thus prevent the development of an abnormal BMI or lipid level, *e.g.*, abnormally high LDL level or abnormally low HDL. The prophylactic methods are similar to therapeutic methods of the present invention and are further discussed in the following subsections.

Therapeutic Methods

The invention further provides methods of treating subjects having a disease or disorder associated with a specific allelic variant of a polymorphic region of an SR-BI gene. Preferred diseases or disorders include those associated with an abnormal body mass, or abnormal lipoprotein (LDL and HDL) levels and disorders resulting therefrom (*e.g.* cardiovascular disease, obesity, cachexia, diabetes, and gallstone formation). In one embodiment, the method comprises (a) determining the identity of the allelic variant; and (b) administering to the subject a compound that compensates for the effect of the specific allelic variant. The polymorphic region can be localized at any location of the gene, *e.g.*, in the promoter (*e.g.*, in a regulatory element of the promoter), in an exon, (*e.g.*, coding region of an exon), in an intron, or at an exon/intron border. Thus, depending on the site of the polymorphism in the SR-BI gene, a subject having a specific variant of the polymorphic region which is associated with a specific disease or condition, can be treated with compounds which specifically compensate for the allelic variant.

In a preferred embodiment, the identity of one or more of the following nucleotides of an SR-BI gene of a subject is determined: nucleotide 146 of exon 1, nucleotide 119 of exon 3, nucleotide 41 of exon 8, nucleotide 54 of intron 5, and nucleotide -41 of intron 10. If a female subject has the more common allele of residue 41 of exon 8 (EX8C), high LDL levels and resulting cardiovascular disorders can be treated, prevented from occurring or can be reduced by administering to the subject a pharmaceutically effective amount of a compound to reduce LDL level to a normal LDL level. Similarly, if a female subject has the less common allele of residue 54 of intron 5 (IVS5T), a high BMI and/or LDL level and consequences thereof, such as diabetes and cardiovascular disorders, can be treated, prevented

from occurring or can be reduced, by administering to the subject a pharmaceutically effective amount of a compound to reduce the BMI and/or the LDL levels. In another embodiment of the invention, if a male subject has the more common allele at residue 41 of exon 8 (EX8C), the more common allele at residue 54 of intron 5 (IVS5C), and the more common allele at residue 146 of exon 1, development of low HDL levels can be treated, prevented or increased by administering to the subject a pharmaceutically effective amount of a compound that increases HDL levels, thereby preventing resulting cardiovascular disorders. Likewise, if a female or a male subject has the more common allele at residue 41 of exon 8 (EX8C), development of low HDL levels and resulting cardiovascular disorders can be treated or prevented from occurring by administering to the subject a pharmaceutically effective amount of a compound which increases HDL level to a normal HDL level. Similarly, if a female subject has the less common allele of residue 54 of intron 5 (IVS5T), low HDL and resulting cardiovascular disorders can be treated or prevented from occurring, by administering to the subject a pharmaceutically effective amount of a compound which increases HDL level to a normal HDL level. Furthermore, if a female subject has both the more common allele at residue 41 of exon 8 (EX8C) and the less common allele of residue 54 of intron 5 (IVS5T), low HDL levels and resulting cardiovascular disorders can be treated or prevented from occurring, by administering to the subject a pharmaceutically effective amount of a compound which increases HDL level to a normal HDL level. In addition, in a female subject, the identification of specific alleles within the SR-B1 gene (*e.g.*, EX8 or IVS5) can be used to predict the effect of HRT on lipid levels (*e.g.*, HDL levels).

Generally, the allelic variant can be a mutant allele, i.e., an allele which when present in one, or preferably two copies, in a subject results in a change in the phenotype of the subject. A mutation can be a substitution, deletion, and/or addition of at least one nucleotide relative to the wild-type allele. Depending on where the mutation is located in the SR-BI gene, the subject can be treated to specifically compensate for the mutation. For example, if the mutation is present in the coding region of the gene and results in an inactive or less active SR-BI protein, the subject can be treated, *e.g.*, by administration to the subject of a nucleic acid encoding a wild-type SR-BI protein, such that the expression of the wild-type SR-BI protein compensates for the endogenous mutated form of the SR-BI protein. Nucleic acids

encoding wild-type human SR-BI protein are set forth in SEQ ID Nos. 1 and 3 and are described, *e.g.*, in Calvo and Vega (1993) J. Biol. Chem. 268:18929.

Furthermore, depending on the site of the mutation in the SR-BI protein and the specific effect on its activity, specific treatments can be designed to compensate for that effect. The SR-BI protein is a cell surface receptor which binds specific forms of lipids, *e.g.*, modified lipid or lipoproteins, *e.g.*, HDL. Thus, an SR-BI protein has an extracellular domain which binds specific molecules, *e.g.*, lipids, a transmembrane domain, and an intracellular domain, which is likely to transmit an intracellular signal. The structure of SR-BI proteins is further described, *e.g.*, in Calvo and Vega, *supra*; Acton et al. (1994) J. Biol. Chem. 269:21003; Acton et al. (1995) Science 271:518; Rigotti et al. (1995) J. Biol. Chem. 270:16221; Fukasawa et al. (1996) Exp. Cell. Res. 222:246; Wang et al. (1996) J. Biol. Chem. 271:21001; and published PCT Application having publication number WO 96/00288 by Acton et al. Thus, if the mutation results in an SR-BI protein which is less capable of binding certain types of modified lipids, resulting in an accumulation of such lipids in the subject, a treatment can be designed which removes such modified lipids from the subject. In one embodiment, a compound which binds this form of lipid and is capable of targeting the lipid to a site where it is eliminated, is administered to the subject. Alternatively, the expression of another cell surface receptor which binds this type of lipid can be increased. In fact, both SR-BI and the class B scavenger receptor CD36 are capable of interacting with anionic phospholipids (Rigotti et al., *supra*). Thus, if a subject has a mutant SR-BI protein which is defective in its binding to anionic phospholipids, the subject can be treated by administration of a compound which increases CD36 protein levels in the cells.

In situations in which the mutant SR-BI protein binds certain forms of lipids with higher affinity, and if this is causing or contributing to a disease, a subject having such a mutated SR-BI protein can be treated, *e.g.*, by administration of compounds which inhibit or decrease the interaction between the specific form of the lipid and SR-BI. For example, soluble forms of SR-BI proteins or binding fragments thereof, can be administered to the subject. Alternatively, small molecules can be administered to the subject for interfering in the interaction between SR-BI and a lipid.

A mutant SR-BI protein can also be an SR-BI protein having a mutation in the cytoplasmic domain of the protein which results in an aberrant signal transduction from the

receptor. Subjects having such a mutation can be treated, *e.g.*, by administration of compounds which induce the same or similar signal transduction or compounds which act downstream of the receptor.

The effect of a mutation in an SR-BI protein can be determined according to methods known in the art. For example, if the mutation is located in the extracellular portion of the protein, one can perform binding assays with specific forms of lipids, *e.g.*, HDL, and determine whether the binding affinity of such lipid with the mutated SR-BI protein is different from the binding affinity of the lipid with the wild-type protein. Such assays can be performed using a soluble form of an SR-BI protein or a membrane bound form of the protein. If the mutation in the SR-BI protein is located in the cytoplasmic domain of the protein, signal transduction experiments can be performed to determine whether the signal transduced from the mutated receptor differs from the signal transduced from the wild-type receptor. Alternatively, one can also investigate whether binding to a protein which interacts with the cytoplasmic domain of the receptor is affected by the mutation. Such determination can be made by, *e.g.*, by immunoprecipitation.

Yet in another embodiment, the invention provides methods for treating a subject having a mutated SR-BI gene, in which the mutation is located in a regulatory region of the gene. Such a regulatory region can be localized in the promoter of the gene, in the 5' or 3' untranslated region of an exon, or in an intron. A mutation in a regulatory region can result in increased production of SR-BI protein, decreased production of SR-BI protein, or production of SR-BI having an aberrant tissue distribution. The effect of a mutation in a regulatory region upon the SR-BI protein can be determined, *e.g.*, by measuring the SR-BI protein level or mRNA level in cells having an SR-BI gene having this mutation and which, normally (i.e., in the absence of the mutation) produce SR-BI protein. The effect of a mutation can also be determined *in vitro*. For example, if the mutation is in the promoter, a reporter construct can be constructed which comprises the mutated promoter linked to a reporter gene, the construct transfected into cells, and comparison of the level of expression of the reporter gene under the control of the mutated promoter and under the control of a wild-type promoter. Such experiments can also be carried out in mice transgenic for the mutated promoter. If the mutation is located in an intron, the effect of the mutation can be determined, *e.g.*, by producing transgenic animals in which the mutated SR-BI gene has been introduced

and in which the wild-type gene may have been knocked out. Comparison of the level of expression of SR-BI in the mice transgenic for the mutant human SR-BI gene with mice transgenic for a wild-type human SR-BI gene will reveal whether the mutation results in increased, decreased synthesis of the SR-BI protein and/or aberrant tissue distribution of SR-BI protein. Such analysis could also be performed in cultured cells, in which the human mutant SR-BI gene is introduced and, *e.g.*, replaces the endogenous wild-type SR-BI gene in the cell. Thus, depending on the effect of the mutation in a regulatory region of an SR-BI gene, a specific treatment can be administered to a subject having such a mutation. Accordingly, if the mutation results in decreased production of an SR-BI protein, the subject can be treated by administration of a compound which increases synthesis, such as by increasing SR-BI gene expression, and wherein the compound acts at a regulatory element different from the one which is mutated. Alternatively, if the mutation results in increased SR-BI protein levels, the subject can be treated by administration of a compound which reduces SR-BI protein production, *e.g.*, by reducing SR-BI gene expression or a compound which inhibits or reduces the activity of SR-BI.

Furthermore, it is likely that subjects having different allelic variants of an SR-BI polymorphic region will respond differently to therapeutic drugs to treat diseases or conditions, such as those associated with an abnormal lipid level. Cholesterol-lowering drugs include lovastatin (MEVACOR; Merck & Co.), simvastatin (ZOCOR; Merck & Co.), dextrothyroxine (CHOLOXIN; Knoll Pharmaceutical Co.), pamaqueside (Pfizer), cholestyramine (QUESTRAN; Bristol-Myers Squibb), colestipol (COLESTID; Pharmacia & Upjohn), acipomox (Pharmacia & Upjohn), fenofibrate (LIPIDIL), gemfibrozil (LOPID; Warner-Lambert), cerivastatin (LIPOBAY; Bayer), fluvastatin (LESCOL; Novartis), atorvastatin (LIPITOR, Warner-Lambert), etofylline clofibrate (DUOLIP; Merckle (Germany)), probucol (LORELCO; Hoechst Marion Roussel), omacor (Pronova (Norway), etofibrate (Merz (Germany), clofibrate (ATROMID-S; Wyeth-Ayerst (AHP)), and niacin (numerous manufacturers). Drugs for treating obesity and/or gallstones include dexfenfluramine (REDUX, Interneuron Pharmaceuticals), megestrol acetate (MEGACE, Bristol-Myers Squibb), Phenylpropanolamine (ACUTRIM; Ciba; and DEXUTRIM; Thompson), fluoxetine (PROZAC, Lilly), dextroamphetamine (DEXEDRINE, SmithKline Beecham), fenfluramine and phentermine, chenodiol (CHENIX, Solvay), orlistat (XENICAL,

Roche), anandamide (Yissum (Israel)), PCM-4 (Omega Pharmaceutical), mono-octanoin (MOCTAN, Stokely-van Camp), sibutramine (MERIDIA, Knoll), testosterone (TESTODERM, Alza), oxandrolone (OXANDRIN, Bio-Technology General), ceruletide diethylamine (TYMTRAN, Pharmacia & Upjohn), testosterone and dihydrotestosterone (ANDROGEL and ANDROGEL-DHT, unimed), somatropin (SEROSTIM, Ares-Serono and BIO-TROPIN, Biotechnology General), and thalidomide (SYNOVIR, Celgene).

A correlation between drug responses and specific alleles of SR-BI can be shown, for example, by clinical studies wherein the response to specific drugs of subjects having different allelic variants of a polymorphic region of an SR-BI gene is compared. Such studies can also be performed using animal models, such as mice having various alleles of human SR-BI genes and in which, *e.g.*, the endogenous SR-BI has been inactivated such as by a knock-out mutation. Test drugs are then administered to the mice having different human SR-BI alleles and the response of the different mice to a specific compound is compared. Accordingly, the invention provides assays for identifying the drug which will be best suited for treating a specific disease or condition in a subject. For example, it will be possible to select drugs which will be devoid of toxicity, or have the lowest level of toxicity possible for treating a subject having a disease or condition.

Computer Readable Means and Arrays

Computer readable media comprising the allelic variants of the present invention is also provided. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. The skilled artisan will readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon an allelic variant of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. Those skilled in the art can readily adopt any of the presently known

methods for recording information on computer readable medium to generate manufactures comprising the allelic variants of the present invention.

A variety of data processor programs and formats can be used to store the allelic variant information of the present invention on computer readable medium. For example, the nucleic acid sequence comprising the allelic variant can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. Any number of dataprocessor structuring formats (*e.g.*, text file or database) may be adapted in order to obtain computer readable medium having recorded thereon the allelic variants of the present invention.

By providing the allelic variants of the invention in computer readable form, one can routinely access the information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

The invention also includes an array comprising allelic variants of the present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 36,000 genes can be simultaneously assayed for expression. This allows a profile to be developed showing a battery of genes specifically expressed in one or more tissues.

In addition to such qualitative determination, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertainable. Thus, genes can be grouped on the basis of their tissue expression *per se* and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression between or among tissues. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat

one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example cardiovascular disorders.

The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells. This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes that could serve as a molecular target for diagnosis or therapeutic intervention.

Other uses for the nucleic acids of the invention

The identification of different alleles of SR-BI can also be useful for identifying an individual among other individuals from the same species. For example, DNA sequences can be used as a fingerprint for detection of different individuals within the same species (Thompson, J. S. and Thompson, eds., Genetics in Medicine, WB Saunders Co., Philadelphia, PA (1991)). This is useful, *e.g.*, in forensic studies.

The present invention is further illustrated by the following examples which should not be construed as limiting in any way. The contents of all cited references (including literature references, issued patents, published patent applications as cited throughout this application are hereby expressly incorporated by reference. The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by

Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No: 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al. eds.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

5. EXAMPLES

The invention now being generally described, it will be more readily understood by reference to the following examples which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention.

Example 1: ***Isolation and sequence analysis of genomic DNA encoding the human SR-BI protein***

A probe consisting of a 474 base pair fragment of the human SR-BI cDNA was used to isolate bacterial artificial chromosomes (BACs) containing genomic DNA encoding the human SR-BI protein from a human BAC library (Research Genetics Inc. (Huntsville, AL) Cat.#96041)). Two BACs were isolated by hybridizing the probe to this library. These BACs were then sized by pulse-field electrophoresis and the inserts were found to be approximately 80 and 70 kilobases long for BAC 179m10 and BAC 256i19, respectively. All further discussion will focus on BAC 179m10.

BAC 179m10 was digested with restriction enzymes and analyzed by Southern blot hybridization with portions of human SR-BI cDNA, and shown to contain a large portion of

the SR-BI sequence. This BAC was then sheared by nebulizing the DNA into fragments of approximately 1-3 kb which were inserted into the pminisk vector and the resulting insert sizes ranged from 1-3 kb. Initially, clones which hybridized to the coding sequence of the full-length human SR-BI cDNA were sequenced, leading to the identification of most of the exons of the gene. Further random sequencing of the BAC sheared library led to the identification of the remaining coding exons and the adjacent intron flanking sequences.

Sequence analysis of the genomic DNA indicated that the human SR-BI gene is at least 50 kb and contains 12 coding exons and one non-coding exon (exon 13, which contains the entire 3' untranslated region). The genomic structure of human SR-BI is shown in Figure 1. The nucleotide sequence of the exons and portions of the introns which are adjacent to the exons is shown in Figure 2A-G. The coding region of the human SR-BI gene consists of 12 exons (see Table IV in the Detailed Description). The location of introns relative to the nucleotide sequence of a cDNA encoding human SR-BI is shown in Figure 2A-G and Figure 3 and indicated in Table V of the Detailed Description. The portions of the protein encoded by each of the exons is also shown in Figure 3 and in Table V.

A number of the introns are extremely large (> 10 kb) (see Table IV in the Detailed Description). The intron/exon boundaries were remarkably similar to those found in the human CD36 gene, which is a member of the same protein family as SR-BI (Tang et al. (1994) J. Biol. Chem. 269:6011).

Example 2: ***Identification of primer pairs to isolate intronic, exonic, and promoter sequences for detection of polymorphisms and mutations***

Multiple pairs of primers were synthesized in order to amplify each of the exons or portions thereof and adjacent intronic regions. Genomic DNA from a human subject was subjected to PCR in 25 µl reactions (1x PCR Amplitaq polymerase buffer, 0.1 mM dNTPs, 0.8 µM 5' primer, 0.8 µM 3' primer, 0.75 units of Amplitaq polymerase, 50 ng genomic DNA) using each of the above described pairs of primers under the following cycle conditions: 94°C for 2 min, 35 x [94°C for 40 sec, annealing temp for 30 sec, 72°C for 1 min], 72°C for 5 min, 4°C hold. The resulting PCR products were analyzed on a 2% agarose gel. The identity of the PCR product was confirmed by digestion with a restriction enzyme and subsequent agarose electrophoresis. Twelve pairs of oligomers were chosen to serve as PCR primers to amplify

regions containing each of the 12 coding exons of the human SR-BI gene and one pair of primers was chosen to serve as PCR primers to amplify a promoter region. The nucleotide sequence of these primers in indicated in Table IX and nucleotide sequences to which these primers bind are shown in Figure 2A-G. The optimum PCR annealing temperature for each primer pair as well as the expected sizes of the PCR products and diagnostic restriction sites is set forth below in Table VIII. Table VIII also indicates the size of DNA fragments obtained when digesting the amplified fragment with the restriction enzyme indicated in the table. A PCR reaction using primers having SEQ ID NO: 41 and 42 for amplifying exon 1 is preferably carried out in the presence of 10% DMSO.

TABLE VII

exon	primer name	SEQ ID NO:	Nucleotide Sequence (5' -> 3')
1	5e16srb1	41	CCCCTGCCGCCGGAATCCTGAAG
	3e16srb1	42	CGCTTTGGCGGAGCAGCCCATGTC
2	5e22srb1	43	TGGGGCCCTCATCACTCTCCTCAC
	3e22srb1	44	GCAGCCTCCCCATCCCGTCCACT
3	5e30srb1	45	ATTGCAGGCGAGTAGAAG
	3e30srb1	46	CAGGCGGGAGGAGAGACA
4	5e41srb1	47	TGGGCTCTTTGCTGTGAGGC
	3e41srb1	48	CCAGGCTGTGTGAGGGGAAG
5	5e50srb1	49	GCCCAGAATGTTCAGACCAG
	3e50srb1	50	GCACCCTCTTCACGACAAAG
6	5e60srb1	51	CACCTGAGAGGGCTTATTA
	3e60srb1	52	CAAAATGCTTTCCAAGTGC
7	5e71srb1	53	GCCGCCGGGTCTGGGTGTCC
	3e71srb1	54	CAGAGGCCAGAGATTAAGCAGAC
8	5e81srb1	55	TTGTATGATGTCCCCTCCCT
	3e81srb1	56	TTCCCACCACCCCAGCCCAC
9	5e91srb1	57	GGTTGACTGTGTCCCTGGAG
	3e91srb1	58	GGGAACACTGGAGCACTGAGC
10	5e104srb1	59	GGTGGTGAGGGTTTAGTGTG
	3e104srb1	60	CTCCCCCCGCCTCCTGCCTC
11	5e112srb1	61	AAGGTGTTGGGTGGCATCTG
	3e112srb1	62	GGCTCCAGGCTGCGGTTGGC
12	5e100srb1	63	TTGAAGAACCGTGTAAAAC
	3e100srb1	64	TTGAGGCTGAAGGAATGA
Prom.	5p13srb1	83	TCCTGGGTGGGCTGGCGAAGTC
	5p13srb1	84	GTTTTGGGGCGGGAGCTGATGAAG

TABLE VIII

Exon	primer pairs	Temp.	Product length	Enzyme check
1	SEQ ID NO: 41	65°C	162 bp	BamHI (144, 118)
	SEQ ID NO: 42			
2	SEQ ID NO: 43	64°C	294 bp	ApaI (189, 98, 7)
	SEQ ID NO: 44			
3	SEQ ID NO: 45	57°C	281 bp	XhoI (153, 128)
	SEQ ID NO: 46			
4	SEQ ID NO: 47	59°C	360 bp	SpeI (292, 68)
	SEQ ID NO: 48			
5	SEQ ID NO: 49	57°C	291 bp	BamHI (157, 134)
	SEQ ID NO: 50			
6	SEQ ID NO: 51	52°C	273 bp	DraII (179, 72, 22)
	SEQ ID NO: 52			
7	SEQ ID NO: 53	59°C	290 bp	EcoRI (184, 106)
	SEQ ID NO: 54			
8	SEQ ID NO: 55	58°C	261 bp	HaeIII (158,103)
	SEQ ID NO: 56			
9	SEQ ID NO: 57	57°C	206 bp	PstI (107,99)
	SEQ ID NO: 58			
10	SEQ ID NO: 59	56°C	253 bp	AvaII (148,105)
	SEQ ID NO: 60			
11	SEQ ID NO: 61	60°C	327 bp	NcoI (242, 85)
	SEQ ID NO: 62			
12	SEQ ID NO: 63	51°C	303 bp	PstI (184,119)
	SEQ ID NO: 64			
prom.	SEQ ID NO: 83	63°C	247 bp	BstXI (200, 47)
	SEQ ID NO: 84			

Example 3: ***Detection of polymorphic regions in the human SR-BI gene by SSCP***

Genomic DNA from a population of 389 unrelated Caucasian men and women, chosen because they had a known HDL and LDL level (high, normal, or low), known body mass index, known level of triglycerides, and known age (see Table IX) was analyzed as described below.

TABLE IX Anthropometric and plasma lipid concentrations of the population studied

	Men (n=101)	Women (n=288)
Age (years)	40 ± 16	36 ± 12
BMI (kg/m^2)	25.2 ± 3.3	22.8 ± 3.6
TC (mg/dL)	227 ± 57	198 ± 45
LDL-C (mg/dL)	158 ± 49	122 ± 39
HDL (mg/dL)	45 ± 23	63 ± 17
TG (mg/dL)	120 ±64	68 ± 34

Genomic DNA from each of these individuals was subjected to PCR in 25 µl reactions (1X PCR Amplitaq polymerase buffer, 0.1 mM dNTPs, 0.8 µM 5' primer, 0.8 µM 3' primer, 0.75 units of Amplitaq polymerase, 50 ng genomic DNA) using each of the above described pairs of primers under the following cycle conditions: 94°C for 2 min, 35 x [94°C for 40 sec, annealing temp for 30 sec, 72°C for 1 min], 72°C 5 min, 4°C hold. The optimum PCR annealing temperatures for each set of primers are given in Table VIII. The expected sizes of the PCR products, as well as diagnostic restriction sites, are also indicated in Table VIII.

The amplified genomic DNA fragments were then analyzed by SSCP (Orita et al. (1989) *PNAS USA* 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; and Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). From each 25 µl PCR reaction, 3 µl was taken and added to 7 µl of loading buffer. The mixture was heated to 94°C for 5 min and then immediately cooled in a slurry of ice-water. 3-4 µl were then loaded on a 10%

polyacrylamide gel containing 10% glycerol and then subjected to electrophoresis either overnight at 4 Watts at room temperature, overnight at 4 Watts at 4°C (for amplifying a promoter region), or for 5 hours at 20 Watts at 4°C (for amplifying exons 3 and 4). The secondary structure of single-stranded nucleic acids varies according to sequence, thus allowing the detection of small differences in nucleic acid sequence between similar nucleic acids. At the end of the electrophoretic period, the DNA was analyzed by gently overlaying a mixture of dyes onto the gel (1x the manufacturer's recommended concentration of SYBR Green I and SYBR Green II in 0.5 X TBE buffer (Molecular Probes)) for 5 min, followed by rinsing in distilled water and detection in a Fluoroimager 575 (Molecular Dynamics). Polymorphisms were found in or near exons 1, 3, 5, 8, and 10.

Example 4: ***Identification of polymorphic regions in the human SR-BI gene by direct sequencing of PCR products***

To determine the sequences of the polymorphisms identified, the regions containing the polymorphisms were reamplified using the aforementioned primers which were modified to contain additional sequence which could be used to directly sequence the PCR product (M13 forward sequence for 5' primer and +M13 reverse sequence for 3' primer) on the 5' end of the primers as listed in Table VII. In particular, the forward primers (5' end primers) contained the nucleotide sequence "TGTAAAACGACGGCCAGT" (SEQ ID NO: 85) located 5' of the nucleotide sequences shown in Table VII and the reverse primer (3' end primer) contained the nucleotide sequence "CAGGAAACAGCTATGACC" (SEQ ID NO: 86) located 5' of the nucleotide sequence shown in Table VI. The genomic DNA from the subjects was subjected to PCR in 50 µl reactions (1x PCR Amplitaq polymerase buffer, 0.1 mM dNTPs, 0.8 µM 5' primer, 0.8 µM 3' primer, 0.75 units of Amplitaq polymerase, 50 ng genomic DNA) using each of the above described pairs of primers under the following cycle conditions: 94°C for 2 min, 35 x [94°C for 40 sec, annealing temp for 30 sec, 72°C for 1 min], 72°C 5 min, 4°C hold. The optimum PCR annealing temperatures for each of the primer pairs are given in Table VIII. The newly amplified products were then purified by agarose gel electrophoresis and subjected to sequencing using M13 forward and reverse primers.

The results indicate that the polymorphism in the region of exon 1 is a change from a guanine at nucleotide 146 to an adenine, resulting in a change of the second amino acid of

the protein from a glycine to a serine. The polymorphism in the region of exon 3 is a change from the guanine at position to 119 to an adenine, resulting in a change of amino acid 135 of the protein from a valine to isoleucine. The polymorphism in the region of exon 8 was determined to constitute a change in base position 41 of exon 8, from a cytidine (referred to herein as EX8C) to a thymidine (referred to herein as EX8T). This substitution does not result in a change in amino acid. In a subpopulation of 142 individuals, about 35% of these individuals were homozygous for an allele having a cytidine at position 41 of exon 8; about 17% of these individuals were homozygous for an allele having a thymidine at this position; and about 48% of these individuals were heterozygous, having one allele of each type. The polymorphism in the region of exon 5 is a change in nucleotide 54 of intron 5 (nucleotide 1 being the first nucleotide of the intron), from a cytidine (referred to herein as IVS5C) to a thymidine (referred to herein as IVS5T). In a subpopulation of 142 individuals, about 24% of individuals have a thymidine at position 54 of intron 5. The polymorphism in the region of exon 11 is a change from the cytidine at position -41 (nucleotide -1 corresponds to the most 3' nucleotide of intron 10) of intron 10 to a guanine. The polymorphisms are indicated in Table VI and in Figure 2A-G.

Example 5: ***Association of common polymorphisms at the SR-BI gene with plasma lipids and anthropometric parameters***

After identification of the mutations in exons 1, 3, and 8 and introns 5 and 10, subjects were typed by digestion of PCR products using the primers and enzymes listed in Table X.

TABLE X Primers and enzymes for typing allelic variants

poly-morphism	primers	temp	digest	product sizes
exon 1 (G/A)	CCGGCGATGGGGCATAAAACCACT (SEQ ID NO.: 89) CGCCCAGCACAGCGCACAGTAGC (SEQ ID NO.:90)	68-62C	AluI	GG: 263 GA: 263, 192, 71 AA: 192, 71
intron 5 (C/T)	GCCCAGAATGTTCAGACCAG (SEQ ID NO.:91) GCACCCTCTTCACGACAAAG (SEQ ID NO:92)	57C	ApaI	CC: 194, 67, 30 CT: 194, 97, 67, 30 TT: 194,97
exon 8 (C/T)	CCTTGTTTCTCTCCCATCCTCACTTCCTCAAGGC (SEQ ID NO.:93) CACCACCCCAGCCCACAGCAGC (SEQ ID NO.:94)	66-61C	HaeIII	CC: 154, 33, 31 CT: 154, 64, 33, 31 TT: 154, 64

Plasma lipids were measured after a 12 to 14-hour overnight fast, using blood collected in tubes containing 0.1% EDTA. Plasma HDL cholesterol was measured after precipitation of plasma apoB-containing lipoproteins as previously described. Plasma total cholesterol, HDL cholesterol and triglyceride levels were measured as previously described. LDL cholesterol was calculated by the Friedewald equation when triglyceride levels were below 400 mg/dL. Coefficients of variation between runs for all lipid assays were less than 5%.

The SPSS statistical package PC version 7.5.1 was used for the statistical analysis. Because of the differences observed between men and women for several of the anthropometric and lipid variables, all the statistical analyses were carried out separately by gender. Triglycerides were log transformed for analysis. For each of the variables examined, the significance of the differences between alleles or genotypes was estimated by analysis of covariance (ANCOVA) using the General Linear Model (GLM) procedure from SPSS, with age as covariate and the Tukey's post hoc test for multiple comparisons for observed means. For analysis related to HDL, smoking and alcohol intake were also included as covariates. Means and standard deviations for all variables were calculated for each genotype group and the significance level was established at $p<0.05$. The allele and haplotype frequencies were estimated using the EH linkage utility program (Terwilliger and Ott J. (1994) Handbook for human genetic linkage. Johns Hopkins University Press, Baltimore, MD).

The frequencies of the less common allele for each of the polymorphisms described at the SR-BI gene locus were as follows: exon 1: 0.1136; exon 3: 0.0184; intron 5: 0.1002; exon 8: 0.4389 and intron 11: 0.0425. The associations between these common polymorphisms and plasma lipid concentrations and BMI are presented in table XI for men and table XII for women. For men, no significant associations were observed between any of the variables examined and the common polymorphisms at exon 1, intron 5 and exon 8. Conversely, in women, the less common allele defined by the polymorphism at exon 8 was associated with significantly lower mean plasma LDL cholesterol concentrations (118±38 and 116±36 mg/dL for heterozygotes and homozygotes, respectively) than those observed in subjects homozygous for the most common allele (131±42 mg/dL; p=0.043). No other significant associations were observed between these polymorphisms and other lipid variables. In women, a significant association was observed between the intron 5 polymorphism and BMI. Women carriers of the less common allele showed a mean BMI value (23.8±3.8) that was significantly greater (p=0.031) than those women homozygous for the most common allele (22.4±3.4).

Table XI Anthropometric characteristics and plasma lipid concentration of the population studied according to SR-BI genotypes (Men)

	Exon 1			Intron 5			Exon 8			
	1/1 (n=71)	1/2 (n=8)	P value	1/1 (n=74)	1/2 (n=19)	P value	1/1 (n=22)	1/2 (n=49)	2/2 (n=18)	P value
Age (years)	41 ± 14	38 ± 16	0.659	40 ± 15	46 ± 13	0.114	47 ± 14	42 ± 14	37 ± 13	0.068
BMI (kg/m^2)	25.7 ± 3.0	25.6 ± 1.9	0.883	25.3 ± 2.6	26.1 ± 3.8	0.326	25.7 ± 3.6	25.8 ± 2.5	25.0 ± 3.2	0.583
TC (mg/dL)	228 ± 46	221 ± 38	0.712	225 ± 55	240 ± 41	0.282	234 ± 40	221 ± 51	225 ± 47	0.559
LDL-C (mg/dL)	159 ± 46	157 ± 47	0.908	156 ± 51	175 ± 48	0.145	168 ± 44	154 ± 52	158 ± 47	0.520
HDL (mg/dL)	43 ± 22	43 ± 14	0.983	43 ± 24	45 ± 19	0.733	39 ± 17	43 ± 25	44 ± 16	0.761
TG (mg/dL)	127 ± 65	105 ± 43	0.346	129 ± 68	98 ± 27	0.052	133 ± 63	122 ± 60	119 ± 69	0.715

Table XII Anthropometric characteristics and plasma lipid concentration of the population studied according to SR-BI genotypes (Women)

	Exon 1				Intron 5			Exon 8			
	1/1 (n=181)	1/2 (n=66)	2/2 (n=1)	P value	1/1 (n=74)	1/2 (n=19)	P value	1/1 (n=73)	1/2 (n=148)	2/2 (n=37)	P value
Age (years)	35 ± 12	38 ± 12	40	0.218	36 ± 12	38 ± 12	0.333	37 ± 12	36 ± 12	34 ± 12	0.458
BMI (kg/m^2)	22.9 ± 3.9	22.5 ± 3.0	19.7	0.552	22.4 ± 3.4	23.8 ± 3.8	0.031	22.8 ± 3.0	23.0 ± 3.9	21 9 ± 3.4	0.252
TC (mg/dL)	197 ± 45	198 ± 42	211	0.932	198 ± 46	204 ± 44	0.446	206 ± 49	196 ± 42	192 ± 42	0.215
LDL-C (mg/dL)	121 ± 39	120 ± 38	134	0.928	122 ± 39	125 ± 42	0.711	131 ± 42	118 ± 38	116 ± 36	0.043
HDL (mg/dL)	63 ± 17	64 ± 19	63	0.930	62 ± 16	64 ± 21	0.145	61 ± 16	64 ± 17	64 ± 22	0.500
TG (mg/dL)	64 ± 28	71 ± 37	72	0.317	68 ± 35	64 ± 30	0.423	67 ± 35	70 ± 38	57 ± 22	0.146

Haplotype analyses (using the method described by Terwillinger and Ott, 1994, *supra*) using the three most common polymorphisms identified at this locus (exon 1, intron 5 and exon 8) were carried out. Six possible haplotypes were identified according to the absence (1) or presence (2) of the variant allele at each one of the three polymorphic sites. Four of the haplotypes were common: 111 (wild type); 112 (exon 8 variant); 121 (intron 5 variant); 211 (exon 1 variant); whereas two of the estimated haplotypes were rare: 221 (variants at exon 1 and intron 5) and 212 (variants at exons 1 and 8). The most common haplotype, 111, had the highest apparent frequency in women (44.6%) and therefore was considered to be wild-type. Each subject was assigned to the most plausible genotype; however, because of the uncertainty associated with genotype assignments in double heterozygotes when studying unrelated subjects on whom the phase of the polymorphisms cannot be directly ascertained, we used in further analysis only those subjects with unequivocal genotypes. In women, the 111/112 and 112/112 genotypes were found to be associated with lower LDL-C levels as compared with the wild genotype (111/111); whereas the 111/121 genotype was associated with increased LDL-C levels as compared with the 111/112 and 111/211 genotypes (Figure 4). In men the trends were similar to those observed in women; however, the differences did not reach statistical significance (Figure 5). In terms of HDL levels, all the alleles carrying mutations were associated with increased HDL levels in men (Figure 7); however, no significant effects were observed in women (Figure 6). Thus, as compared to the wild-type haplotype, the analyses revealed that men with haplotypes 112, 121, and 211 tended to have significantly higher HDL. In fact, those subjects with haplotype 211 (one wild-type chromosome and one with a polymorphism in exon 1) had an average of 75% higher HDL levels than individuals containing only wild-type chromosomes. Such associations were not observed in women. Without wanting to be limited by a specific mechanism of action, it has been shown that, though less efficient than the LDL receptor, SR-BI is able to mediate the degradation of LDL *in vitro*. SR-BI may also play an indirect role in LDL cholesterol metabolism by altering cholesterol homeostasis in the individual.

The previous association between the intron 5 polymorphism and BMI observed in women using single marker analysis was stronger using haplotype analysis. Subjects carrying the 121 haplotype (n=12), had a mean BMI (25.0"3.3) significantly greater ($p<0.05$) than that observed in subjects homozygous for the wild haplotype 111 (n=29; 22.8"3.2) and those

carriers of the 112 (n=78; 22.8"3.9) and 211 (n=12; 21.5"2.1) haplotypes (Figure 8). However, no association was observed in men (Figure 9).

There was also a significant association between the intron 5 polymorphism and BMI in women. The latter finding was most significant in premenopausal women. The 288 female subjects in this study were not considered to be obese. Thus, this effect was observed on individuals within the normal weight range. There are only a few polymorphisms known to date that are associated with BMI values in humans (Bouchard and Perusse (1996) *Obesity Research* 4:81-90), and thus this is a significant finding. Increased body-mass index has been associated with higher mortality from all causes and from cardiovascular disease. For mortality from cardiovascular disease, the relative risk associated with an increment of one in the body-mass index in women in the age range of 30-to-44 year old has been reported to be 1.08 (95 percent confidence interval, 1.05 to 1.11) (Stevens et al. (1998) *N Engl J Med* 338:1-7; Colditz et al. (1995) *Ann Intern Med.* 122(7):481-486). The data presented herein show that for an average woman, the presence of the 121 haplotype raises the BMI by approximately 2.2 kg/m2 which corresponds to about 6 kgs. This increase in BMI could result in an increase in CHD mortality of about 17.6%, primarily due to a greater risk of developing non-insulin dependent diabetes mellitus (Stevens et al. (1998) *supra*; Colditz et al. (1995) *supra*), a major risk for coronary artery atherosclerosis. An intriguing observation is the fact that no subjects were found to be homozygous for this polymorphism in this sample of 389 randomly selected individuals, despite the expected frequency of 1% (~4 individuals) assuming Hardy-Weinberg equilibrium.

Example 6: ***<u>Association of common polymorphisms at the SR-BI gene with plasma lipids in three ethnically distinct populations</u>***

<u>Study populations</u>

Subjects were drawn from three collections of nuclear families ascertained for T2DM. Two Scandinavian populations were recruited: one from Finland and one from southern Sweden. The third was a population of Israelis of Ashkenazi Jewish origin. Nuclear families were identified having at least two members with T2DM. Those patients with a diagnosis before the age of 70 were invited to participate. In the Ashkenazi study,

at least one sibling had to have initial diagnosis prior to age 56. All subjects with age of onset <35 years were excluded to minimize the inclusion of Maturity Onset of the Young (MODY) and Type 1 diabetes mellitus. Furthermore, patients who became insulin-dependent within two years of diagnosis were excluded. To avoid bilinial inheritance, families in which both parents were known to have diabetes were excluded.

T2DM was diagnosed using WHO criteria (World Health Organization Study Group Diabetes Mellitus (1985) Technical report series No. 727, WHO, Geneva), *e.g.,* fasting blood glucose >6.7 mmol/l or two hour blood glucose ≥10.0 mmol/l. Individuals lacking fasting blood glucose and oral glucose tolerance test data were considered affected if currently taking oral hypoglycemics and/or insulin. To test unaffected status, all presumed unaffected siblings were asked to undergo a standard oral glucose tolerance test, however, inclusion in the study was not dependent on their agreement to undergo this test. These studies were approved by the local Institutional Review Boards in Israel, Finland and Sweden. Informed consent was obtained from all participants.

Blood samples were obtained following an overnight fast of at least 12 hours by drawing into tubes containing EDTA. Plasma HDL, total cholesterol and TG were measured using standard procedures. Clinical data, including anthropometric measures such as body mass index (BMI; weight in kilograms/height in meters squared), were gathered for each subject on the day of enrollment.

SRB1 variants

The SR-B1 variants evaluated were chosen based on allele frequencies to maximize the ability to detect an association. The most common DNA variants described in SR-B1 are a silent single nucleotide polymorphism (SNP) in exon 8 at amino acid 350, and a SNP 54 nucleotides into intron 5 (IVS5+54C→T). The SR-B1 variants in exon 8 (EX8; alleles EX8C and EX8T) and intron 5 (IVS5; alleles IVS5C and IVS5T) were examined to determine if these two SNPs alone or in combination were associated with HDL levels.

Genotyping

Genomic DNA was isolated from peripheral blood lymphocytes using the Puregene kit (Gentra Systems, Inc.) according to manufacturer's suggested protocol. The IVS5 and EX8 polymorphisms were amplified under the following PCR conditions: 2.5 mM MgCl2, 1X Ammonium Sulfate buffer (80 mM (NH4)2SO4, 335 mM Tris-HCl pH 8.8, 0.05% Tween 20), 200uM dNTP, 0.015 U Platinum *Taq* polymerase (Life Technologies, Inc.), 0.17 uM each of forward and reverse primers, and 5 ng genomic DNA in a 15 uL reaction volume. Forward PCR primers, shown in Table X, were fluorescently labeled at the 5' end. Thermocycling conditions were 95°C for 10 min, 95°C for 30 s, 55°C for 30 s, 72°C for 1 min, for 35 cycles. Variants were visualized by Single-Stranded Conformation Polymorphism (SSCP) on ABI Prism 377 DNA Sequencers (Applied Biosystems Group) by loading products on a non-denaturing 7% Long Ranger acrylamide gel (BioWittiker Molecular Applications) at 75 W, 15°C, for 10 h.

TABLE XIII. Sequence of primers used to genotype SR-B1 variants in intron 5 and exon 8.

Polymorphism	Primer sequence	Product size (bp)	Location on SSCP gel	Allele
IVS5 (C/T)	TCACGGGGGTCCAGAACATC (SEQ ID NO: 122)	174	upper band	T
	TTCACGACAAAGGAAGAAGGAGC (SEQ ID NO:123)		lower band	C
EX8 (C/T)	TGTCGGGTATTATGGTCATCGCC (SEQ ID NO:124)	238	upper band	T
	ATGTCCACGAACAAGGAGTGTGC (SEQ ID NO:125)		lower band	C

Design of nested case-control studies

Sex-specific thresholds were used for defining low HDL cases and high HDL controls. Cases and controls were chosen without respect to T2DM status. For women, low HDL was defined as having HDL ≤1.16 mmol/l (45 mg/dl), representing the median

HDL level in women from all three populations combined. Female controls had HDL >1.16 mmol/l. For men, low HDL was defined as having HDL ≤ 1.01 mmol/l (39 mg/dl), representing the median HDL level in men from all three populations combined. Male controls had HDL >1.01 mmol/l.

Inclusion of multiple family members in either case series or control series may affect the distribution of genotypes and lead to an overestimate of the significance of the association. Therefore, for each sex an independent case series and independent control series was defined. If more than one case existed in the same family, the case with the younger age was selected to remain in the case series. If more than one control existed from the same family, the control with the younger age was selected to remain in the control series. Younger aged controls were chosen to improve matching on age-related covariates. If both a case and control were found in the same family, the case was used and the control eliminated.

Statistical analysis

All analyses were performed using the SAS statistical package version 6.12 (SAS Institute Inc., Cary, NC). Differences in mean levels of HDL between genotype groups was compared using one-way analysis of variance (ANOVA). A T-Test was used to assess differences between cases and controls with respect to age, BMI and TG. Since TG levels had a skewed distribution, the statistical analyses were based on log-transformed data. However, in the tables the TG levels are given as mean (±SD). The chi square statistic was used to compare the proportion of T2DM among cases and controls. Odds ratios and 95% confidence intervals were calculated for contingency tables and the overall significance assessed using the continuity adjusted chi square. Logistic regression analysis was performed using the PROC LOGISTIC procedure in SAS. BMI, age and TG were entered into the models as continuous variables while T2DM status was dichotomous. Linkage disequilibrium was measured using the normalized disequilibrium parameter (Lewontin, RC (1964) *Genetics* 49:49-67), D', and the statistical significance assessed

with a chi square test. Relative risk and attributable risk estimates were calculated as described below.

Calculation of relative risk and attributable risk from case-control data.

RR= relative risk; OR= odds ratio; D= outcome; G= risk genotype.

In this study, p(D), the frequency of low HDL, is ~50% as it was defined as being below the population median.

The following conditional probabilities can be estimated from the two by two table of case-control data:

p(G|D)	p(G|noD)
p(noG|D)	p(noG|noD)

Since p(D) is known, P(G) can be calculated using the law of total probability:

$$P(G) = p(G|D)\,p(D) + p(G|noD)\,p(noD)$$

Relative risk and odds ratio estimates are related in the following way:

$$RR=R_1/R_0 \qquad OR=RR\,(1-R_0)/(1-R_1) \qquad RR=OR(1-R_1)/(1-R_0)$$

R_1 and R_0 can be expressed as conditional probabilities:

$R_1 = p(D\|G)$	$R_0 = p(D\|noG)$
$(1-R_1) = p(noD\|G)$	$(1-R_0) = p(noD\|noG)$

Using Bayes theorem, these conditional probabilities can be expressed as:

$p(noD|G) = p(G|noD)\,p(noD)/p(G)$ $\qquad$ $p(noD|noG) = p(noG|noD)\,p(noD)/p(noG)$

Therefore: $(1-R_1)/(1-R_0) = [p(G|noD)\,p(noD)/p(G)]/[p(noG|noD)\,p(noD)/p(noG)]$

and relative risk can be estimated as:

$$RR = OR\,[p(noG)\,p(G|noD)/p(G)\,p(noG|noD)]$$

Attributable risk is calculated from the following formula described by Coughlin *et al*[26].

$$AR = 1 - \frac{\Sigma (n_j/R_j)}{N}$$

where *j* are the strata defined by presence or absence of the genotype, n_j is the number of cases in each stratum and R_j is the relative risk associated with each stratum. N is the total number of cases.

Results

Mean HDL levels were correlated with SR-B1 EX8 genotype for all family members with genotype data available (1649 individuals representing 600 families from three populations). All three populations suggested an association between SR-B1 genotype and HDL level. Specifically, carriers of EX8C have lower HDL than non-carriers. Two of the populations, the Ashkenazi and Finnish, showed marked differences in this effect by sex with the effect being more pronounced in women. To further characterize the association of SR-B1 variants with HDL, nested sex-specific case control studies were designed for each population and sex. Table XIV, below, shows the breakdown of cases and controls and their sampling from the initial collection of families. The final sample size included 558 low HDL cases and 379 high HDL controls. Each population and sex was examined separately to determine the association of SR-B1 EX8 and IVS5 genotypes with low HDL.

TABLE XIV. Breakdown of individuals chosen for nested case-control studies.

	WOMEN			MEN		
	AKZ	FIN	SWE	AKZ	FIN	SWE
all individuals genotyped:	318	346	221	266	273	225
all cases (low HDL):	171	162	89	152	110	100
all controls (high HDL):	147	184	132	114	163	125
independent cases:	121	108	65	113	78	73
independent controls:	114	110	89	87	114	92
case & control in same family:	32	61	33	23	42	46
final cases	121	108	65	113	78	73
final controls:	82	49	56	64	72	56

Characteristics of the six population and sex-specific pairs of cases and controls are shown in Table XV, below. The average age of study participants varied between 56 and 63 years. Significant differences in age between cases and controls were found for Ashkenazi women ($p<.0006$) and Swedish men ($p<.01$). Because the cases and controls were chosen from families originally ascertained for T2DM, they are enriched for T2DM and related phenotypes. Most (80%) of cases and controls combined had T2DM. However, significant differences in the frequency of T2DM between cases and controls were found only for Finnish women and men (both $p<.01$). Significant differences in mean BMI between cases and controls were found only for Swedish women ($p<.01$). All populations were enriched for low HDL and high TG, the hallmark dyslipidemia associated with T2DM. HDL and TG levels are closely linked and reflected in the fact that all six pairs of cases and controls, defined as such based on HDL levels, also differed significantly in their levels of TG.

TABLE XV. Characteristics of cases and controls by sex and population.

	WOMEN		MEN	
	low HDL	high HDL	low HDL	high HDL
ASHKENAZI				
No. individuals	121	82	113	64
% T2DM	75%	85%	82%	91%
mean age	57.9±11.5	63.0±9.4 *	57.7±12.3	59.8±10.2
mean BMI	28.7±4.528.4±4.8	27.3±3.727.4±4.0		
mean HDL (mmol/l)	0.89±.211.46±.27 †		0.76±.191.28±.34 †	
mean TG (mmol/l)	2.37±1.41	1.71±.76 *	2.41±1.87	1.69±1.00 *
FINNISH				
No. individuals	108	49	78	72
% T2DM	84%	59% ‡	72%	90% ‡
mean age	59.7±11.9	58.4±13.1	56.4±11.9	57.5±11.5
mean BMI	30.5±5.228.5±5.8		31.8±17.1	27.9±3.9
mean HDL (mmol/l)	1.00±.131.50±.23 †		0.87±.101.32±.32 †	
mean TG (mmol/l)	1.97±.821.26±.64 †		2.19±1.08	1.42±.76 †
SWEDISH				
No. individuals	65	56	73	56
% T2DM	82%	73%	79%	79%
mean age	58.4±13.4	58.2±13.0	56.1±13.5	61.4±9.8 ‡
mean BMI	29.4±5.626.2±4.6 ‡		27.8±3.926.3±4.0	
mean HDL (mmol/l)	0.93±.151.51±.29 †		0.85±.121.31±.24 †	
mean TG (mmol/l)	2.14±1.06	1.36±.67 †	2.40±1.82	1.39±.62 †

‡p<.01; *p<.001; †p<.0001

The association of each of the two SR-B1 variants with HDL was assessed separately in both univariate and multivariate analyses. Unadjusted and adjusted odds ratios, taking into account T2DM, age, BMI and TG levels, for the association of SR-B1 EX8 and IVS5 genotypes with low HDL are presented in Table I for women and Table II for men, in the Detailed Description. Univariate analyses demonstrated consistent and

strong associations between presence of EX8C and low HDL in women from all three ethnic groups. Unadjusted odds ratios were 2.59, 2.92 and 2.33 for the Ashkenazi, Finnish and Swedish women, respectively. All populations achieved statistical significance although the Swedish population was borderline (p=.054). In men, carriers of EX8C also demonstrated increased odds of having low HDL, although the effect was not as strong as in women. Unadjusted odds ratios were 1.95, 1.38 and 2.82 for the Ashkenazi, Finnish and Swedish populations, respectively. Only the Finnish population failed to reach statistical significance (p=.50). Odds ratios are only estimates of the underlying relative risk and should therefore be interpreted cautiously. Whereas the combined population unadjusted odds ratios for EX8C carriers were 2.66 (p<.00001) for women and 1.81 (p=.008) for men, the relative risks (calculated using the formulas set forth below) are more modest; 1.74 for women and 1.40 for men.

Odds ratios for the association between carriers of IVS5T and low HDL were consistently elevated across all three ethnic groups; 2.06, 4.95 and 3.34 for Ashkenazi, Finnish and Swedish women, respectively. Only the Swedish reached statistical significance (p=.013). For men, presence of IVS5T had the opposite effect, being inversely associated with low HDL. Odds ratios were 0.97, 0.37 and 0.30 for Ashkenazi, Finnish and Swedish men, respectively. Only the Swedish reached statistical significance (p=.024).

Multivariate logistic regression analysis was performed to control for the possible confounding effects of T2DM, age, BMI and TG on the association of SR-B1 genotypes with HDL. Inclusion of these terms as covariates in the logistic regression models had no marked effect on the association between SR-B1 variants and low HDL. Adjusted odds ratios are presented in Tables I and II in the Detailed Description. SR-B1 genotypes remained as strong, independent predictors of low HDL. An additional logistic regression model was run to test for the interaction of ethnic group with SR-B1 exon 8 genotype. It revealed no significant interaction in either men or women. Therefore data on all three ethnic groups were combined. Adjusted odds ratios for the association of SR-B1 genotypes with HDL in the combined populations were 2.84 (p<.0001) and 2.32 (p=.01) in

women for EX8C and IVS5T, respectively. In men, the adjusted odds ratios were 1.79 (p=.015) and 0.64 (p=.15) for EX8C and IVS5T, respectively.

Both EX8C and IVS5T alone were positively associated with low HDL in women. In men, EX8C appears to have a positive association with low HDL while IVS5T shows an inverse, borderline significant association. EX8C and IVS5T are in complete positive linkage disequilibrium in the Ashkenazi and Finnish populations (D'=1.0, $p<10^{-6}$) and near complete linkage disequilibrium in the Swedish population (D'=.91, $p<3x10^{-6}$). The combined effect of the two variants on low HDL are shown in Table III. In the Detailed Description. Among women, carriers of both variants had the highest odds of having low HDL (OR=4.79, p<.00001). However, even carriers of EX8C alone were at increased odds of having low HDL relative to those without either variant (OR=2.44, p<.0001).

The combination of variants at EX8 and IVS5 yielded somewhat different results in men. As in women, male carriers of EX8C who lacked IVS5T were at increased odds of having low HDL (OR=1.95, p<.01). However, unlike women, male carriers of both EX8C and IVS5T were no more likely to have low HDL than individuals with neither variant (OR=1.08, p=.95). This difference was not due to population (ethnic) differences, as consistent associations are noted across the ethnic groups.

Attributable risk estimates were calculated from the data for men and women separately to determine the portion of low HDL which may be attributable to having the SR-B1 EX8. By definition, p(D), the frequency of low HDL in families of T2DM patients, was about 50% since the cutoffs for choosing cases and controls were based on population medians. Using the formulas presented above (see "Calculation of relative risk and attributable risk from case-control data"), risk of low HDL attributable to carrying SR-B1 EX8C was calculated to be 35% for women and 23% for men in this population.

All of the above-cited references, patents and publications are hereby incorporated by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.